 doi:10.20944/preprints201906.0072.v1

Peer-reviewed version available at *Risks* **2019**, *7*, 80; doi:10.3390/risks7030080

Article

Quantile Regression with Telematics Information to Assess the Risk of Driving Above the Posted Speed Limit

Ana M. Pérez-Marín [1], Montserrat Guillen [1,*], Manuela Alcañiz [1] and Lluís Bermúdez [2]

[1] Dept. Econometrics, Riskcenter-IREA, Universitat de Barcelona, Av. Diagonal, 690, 08034 Barcelona, Spain; amperez@ub.edu (A.M.P.-M.); malcaniz@ub.edu (M.A.)
[2] Dept. Matemàtica Econòmica, Financera i Actuarial, Universitat de Barcelona, Av. Diagonal, 690, 08034 Barcelona, Spain; lbermudez@ub.edu
* Correspondence: mguillen@ub.edu; Tel.: +34-93-403-7039

Abstract: We analyze real telematics information for a sample of drivers with usage-based insurance policies. We examine the statistical distribution of distance driven above the posted speed limit – which presents a strong positive asymmetry – using quantile regression models. We find that, at different percentile levels, the distance driven at speeds above the posted limit depends on total distance driven and, more generally, on such factors as the percentages of urban and nighttime driving and on the driver's gender. However, the impact of these covariates differs according to the percentile level. We stress the importance of understanding telematics information, which should not be limited to simply characterizing average drivers, but can be useful for signaling dangerous driving by predicting quantiles associated with specific driver characteristics. We conclude that the risk of driving long distances above the speed limit is heterogeneous and, moreover, we show that prevention campaigns should target primarily male, non-urban drivers, especially if they present a high percentage of nighttime driving.

Keywords: telematics; motor insurance; speed control; accident prevention

---

1. Objective

Every kilometer driven above the posted speed limit increases the risk of accident. This is the hazard to which the driver, the passengers in the vehicle and those in vehicles on the same stretch of road expose themselves. The main objective of this paper is to analyze, in a real case telematics data set, the distribution of the distance traveled at speeds above posted limits and to show that it is dependent on the total distance driven and other factors that include the percentages of urban and nighttime driving and the driver's gender. If we only model the mathematical expectation, i.e. the average distance driven at speeds above the posted limits, significant relationships are likely to be found with a number of telematics covariates. However, here, we consider quantile regression to determine whether the impact of certain factors might differ depending on the percentile being analyzed.

When quantile regression slopes differ depending on the level, the risk of driving above the posted speed limit is not homogeneous across all drivers, begging the question as to how this risk might be predicted or measured. Thus, in this paper, we also seek to show how specific driver characteristics can help predict a driver's expected ranking, that is, not in relation to the whole population, but to similar drivers.

The rest of this paper is organized as follows. In section 2, we present the background to this study. In section 3, the theory of quantile regression modelling and the data set used in this study are presented. In section 4, the results are discussed and, finally, in section 5, we outline the conclusions that can be drawn.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2019 doi:10.20944/preprints201906.0072.v1

Peer-reviewed version available at Risks 2019, 7, 80; doi:10.3390/risks7030080

2. Background

There is much evidence in the literature pointing to the relationship between elevated vehicle speeds and the risk of collision (see Ossiander and Cummings, 2002, and Vermon et al., 2004, among others). Likewise, the effectiveness of speed cameras in the reduction of road traffic collisions and related casualties has been extensively demonstrated (see Pilkington and Kinra, 2005, and Wilson et al., 2007, among others), which would seem to confirm that high speeds increase the risk of collision. Speeding, moreover, has been shown to be directly related to the severity of accidents (see, among others, Dissanayake and Lu, 2002, and Jun et al., 2007, 2011), while Yu and Abdel-Aty (2014) report that marked variations in speed prior to a crash increase the likelihood of severe accidents.

Not all drivers present the same tendency to exceed the posted speed limit. More specifically, evidence of gender differences in driving patterns has been reported in many articles (see Ayuso et al., 2014, 2016a, and 2016b). It has been shown that, compared to women, men present riskier driving behavior, driving more kilometers per day, during the night and at speeds above the limit. All these factors have been shown to be related to a greater number of accidents (Gao et al. 2019, Gao and Wüthrich, 2019 and Guillen et al. 2019). For example, Paefgen et al. (2014) found that the risk of accident is higher at nightfall, during the weekends on urban roads and at low-range (0-30 km/h) or high-range speeds (90-120 km/h).

Speed control has recently come under investigation in connection with advanced driver assistance systems (ADAS) and semi-autonomous vehicles. Pérez-Marín and Guillen (2019), for example, analyzed the contribution of telematics information and usage-based insurance (UBI) research in identifying the effect of driving patterns – above all, speeding – on the risk of accident. The authors used a predictive model of the number of claims in a portfolio of insureds as their starting point for addressing risk quantification in relation to vehicles exceeding the speed limit. They concluded that if excess speeds could be eliminated, the expected number of accident claims could be reduced by half, in the average conditions prevailing in their real UBI dataset. Pérez-Marín et al. (2019) show that young drivers tend to reduce posted speed limit violations after an accident.

It has also been demonstrated that both the mean speed and the coefficient of variation of speed are relevant risk factors (Taylor et al., 2002). Moreover, interest has been expressed in the percentile assessment of the speed distribution, as opposed to just the mean. In this regard, Hewson (2008) claims that controlling the 85th percentile speed is common when designing road safety interventions. The same author also examined the role of quantile regression for modelling this percentile and specifically demonstrated its potential benefits when evaluating whether or not an intervention is able to significantly modify the 85th percentile speed.

Hewson (2008) based his analysis on a data set of observations on approximately 100 vehicle speeds at each of 14 pairs of sites recorded before, right after and some time after the intervention (the installation of warning signs, in this instance). However, here, we apply quantile regression to an analysis of the effects of telematics information on a range of percentiles of the distance travelled at speeds above the limit, rather than to the speed measured at one specific moment in time.

We should stress that the objective of our paper is not the same as Hewson's (2008), inasmuch as we do not seek to evaluate a particular safety intervention. Our aim is to understand conditional quantiles of distance traveled, possibly at different moments, rather than an instant speed measurement. To do so, our analysis is based on real telematics information from a sample of drivers covered by a UBI policy. This means that, in addition to speed, we analyze other telematics variables, such as the location and time of driving and the total distance travelled by each driver in the sample.

3. Methods

3.1. Quantile regression

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2019 doi:10.20944/preprints201906.0072.v1

Peer-reviewed version available at Risks 2019, 7, 80; doi:10.3390/risks7030080

Our quantile regression model follows the same notation as that used in Hewson (2008). Thus, in the classical multiple linear regression model, the response y is modeled as

$$y_i = x_i^T \beta + \epsilon_i$$

where $x_i = (1, x_{i1}, \dots, x_{ip})$, in which p is the number of explanatory variables, β is the vector of coefficients such that $\beta = (\beta_0, \beta_1, \dots, \beta_p)$ and ϵ is the random term with distribution $N(0,\sigma^2)$. When we model the conditional mean response, the Gaussian likelihood function is given by

$$L(\beta) \propto exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - x_i^T\beta\right)^2\right\}.$$

The least squares estimation of β is obtained by maximizing $L(\beta)$ over β.

As we aim to estimate a conditional quantile function 100α%, rather than a conditional mean, we need to use a quantile regression model (see Koenker and Hallock, 2001, and Yu et al., 2003, among others). The objective function to be minimized in this case equals

$$L_\alpha(\beta) \propto exp\{-\textstyle\sum_{i=1}^{n}\rho_\alpha\left(y_i - x_i^T\beta\right)\},$$

where the expression contains an asymmetric loss function ρ_α. To explain just what this asymmetric loss function is, we need to introduce some notation. We consider that the 100α% quantile of the residual ϵ is the 100α% largest value (that is, it has 100α% of values smaller than it and 100(1-α)% of values larger than it). Quantile regression, therefore, involves finding estimates $\hat{\beta}$ where 100α% of the residuals are below zero, and 100(1-α)% are above zero. We use an indicator function I_A on the set A, as

$$I_A(\delta) = \begin{cases} 1 & \delta \in A \\ 0 & \delta \notin A \end{cases}$$

The loss function ρ_α can then be defined as follows:

$$\rho_\alpha(\delta) = \alpha \ \epsilon \ I_{(-\infty,0]}(\delta) - (1-\alpha)\ \epsilon \ I_{[0,\infty)}(\delta) \qquad (1)$$

for any value of α between 0 and 1. Finding the values of $\hat{\beta}$ that maximize the likelihood of the quantile regression model is the same as finding the values of $\hat{\beta}$ that minimize this loss function (see Hewson, 2008). Equation (1) can be minimized by using linear programming techniques. The function qr of the quantreg R package (Koenker et al., 2018) can be used to fit a quantile regression model.

3.2. *The data*

The data set comprises a sample of 9,614 drivers with UBI coverage, which targets drivers between the ages of 18 and 35, for the whole of 2010. The variables are presented in Table 1. Age is the age of the driver at the beginning of 2010. We also have information on gender (Gender), total number of kilometers (km) driven during 2010 (Km) and its natural logarithm (Lnkm). Note that we considered the natural logarithm of Km, Lnkm, as it has been shown that distance travelled has a nonlinear effect on the risk of an accident (see Boucher et al., 2013). We also have information on the number of kilometers driven at speeds above the posted limit (Tolerkm, which is the dependent variable), percentage of km driven on urban roads (Porc_vurba) and, finally, percentage of kilometers driven at night (Porc_nocturn). All the drivers had UBI coverage throughout the whole of 2010 and all the telematics variables refer to this year.

Table 1. Variable description.

Variable	Description

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2019 doi:10.20944/preprints201906.0072.v1

Peer-reviewed version available at *Risks* **2019**, *7*, 80; doi:10.3390/risks7030080

Tolerkm	Number of kilometers driven at speeds above the posted limit during 2010.
Km	Total number of kilometers driven during 2010.
Lnkm	Logarithm of the total number of kilometers driven during 2010.
Porc_vurba	% of kilometers driven on urban roads during 2010.
Porc_nocturn	% of kilometers driven at night (between midnight and 6 am.) during 2010.
Age	Age of the driver at the beginning of 2010.
Gender	1 = Male, 0 = Female

Table 2. Descriptive statistics.

Variable	Min	1st Qu	Median	Mean	3rd Qu	Max	St. Dev.	Skewness
Tolerkm	0.00	282.40	689.20	1,398.20	1,701.60	23,500.20	1,995.37	3.64
Km	0.69	7,530.56	11,697.82	13,063.71	17,337.00	57,756.98	7,715.80	1.08
Lnkm	-0.37	8.93	9.37	9.27	9.76	10.96	0.75	-1.87
Porc vurba	0.00	15.60	23.39	26.29	34.32	100.00	14.18	1.03
Porc_nocturn	0.00	2.48	5.31	7.02	9.84	78.56	6.13	1.67
Age	18.11	22.66	24.63	24.78	26.88	35.00	2.82	0.11

The gender distribution of the sample is 49% women and 51% men. Table 2 shows that the average age of drivers in the sample is 24.78 years. The average number of kilometers travelled during the year was 13,063.71 (standard deviation of 7,715.80). We also observe that on average drivers travel 26.29% of kilometers on urban roads and 7.02% of kilometers at night. The mean of kilometers travelled at speeds above the limit (Tolerkm, dependent variable) is 1,398.20, while its median is 689.20. Tolerkm has positive asymmetry (skewness coefficient equals 3.64); the distribution has a long tail as can be observed in Figure 1. The rest of the variables also present some degree of skewness, but not as high as Tolerkm.

4. Results

We fitted a multiple linear regression model to the variable Tolerkm, although we consider it unsuitable insofar as the dependent variable is highly asymmetric. The variable Km was included in the model as its natural logarithm (variable Lnkm), as it produced a better fit. Parameter estimates are shown in Table 3. The R-squared goodness-of-fit statistic equals 0.26.

All the explanatory variables have a significant effect except for Age, which is attributable to the fact that UBI policies were sold primarily to young drivers and, so, the age range in the sample is not wide. Lnkm and Porc_nocturn present positive parameter estimates, indicating that increases in the total number of kilometers driven and in the percentage of km driven at night contribute to increase the expected number of kilometers driven at speeds above the posted limits. Porc_vurba, in contrast, has the opposite effect, the higher the percentage of kilometers driven on urban roads, the lower the expected number of kilometers driven at speeds above the posted limit. Finally, gender (indicating males) has a positive parameter estimate, meaning that, on average, men drive more kilometers at speeds above the posted limit than women.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2019 doi:10.20944/preprints201906.0072.v1

Peer-reviewed version available at *Risks* **2019**, *7*, 80; doi:10.3390/risks7030080

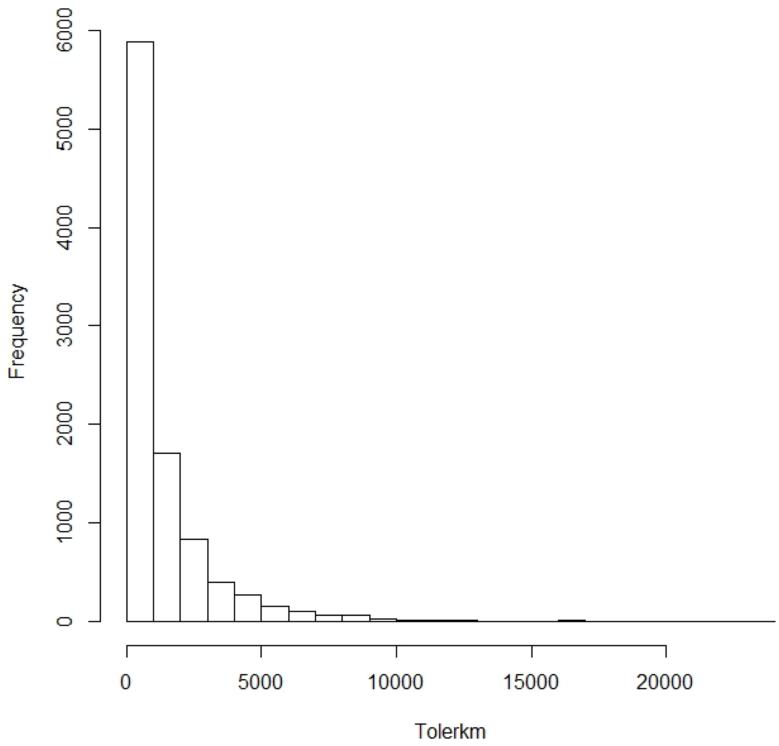

Figure 1. Histogram of the distance travelled at speeds above the limits.

Table 3. Parameter estimates of the linear regression model.

	Parameter estimate (p-value)
Intercept	-8082.506 (<0.0001)
Lnkm	1064.506 (<0.0001)
Porc_vurba	-21.868 (<0.0001)
Porc_nocturn	7.536 (0.0101)
Age	-1.131 (0.8565)
Gender	328.009 (<0.0001)

To fulfil the objectives identified in the first section and, at the same time, to address the strong positive asymmetry, a grid of quantile regressions with different percentiles were fitted to the data. The results of the quantile regression models are presented in Table 4. Each column shows the parameter estimates of the quantile regression at the following percentiles: 50th, 75th, 90th, 95th, 97.5th and 99th. In general, significant parameter estimates are the same as those found in the multiple linear regression model shown in Table 3. However, the results in Table 4 show that the covariates have different marginal effects on conditional quantiles depending on the estimated percentile. These changes in the parameters depending on the quantile level at which the model is specified are clearly illustrated in Figure 2 and are discussed in detail below.

Table 4. Parameter estimates of the quantile regression model for different percentiles.

	50th percentile (p-value)	75th percentile (p-value)	90th percentile (p-value)	95th percentile (p-value)	97.5th percentile (p-value)	99th percentile (p-value)
Intercept	-4496.53	-6250.34	-6418.11	-6009.63	-5137.24	-2451.17
	(<0.0001)	(<0.0001)	(<0.0001)	(<0.001)	(<0.0001)	0.5780
Lnkm	597.60	892.80	1074.66	1094.57	1119.94	1180.21
	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.001)
Porc_vurba	-9.19	-22.26	-39.59	-53.44	-68.58	-87.12
	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)
Porc_nocturn	5.41	6.71	21.76	37.49	20.01	43.86
	(<0.0001)	(0.0363)	(0.0226)	(0.0086)	(0.4266)	(0.4014)
Age	-2.56	1.84	5.16	40.29	71.28	36.87
	(0.1632)	(0.7298)	(0.7419)	(0.2086)	(0.1094)	(0.7009)
Gender	206.76	377.94	574.08	755.87	1070.06	1091.38
	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(<0.0001)	(0.0624)

First, Table 4 shows that the percentage of kilometers driven at night presents a highly significant effect when we estimate the 50^{th} percentile and that it remains significant – at the 5% level – but with a larger p-value, when we estimate the 75^{th} 90^{th} and 95^{th} percentiles. Likewise, the effect of gender is positive and significant at the 5% significance level for all quantiles, except for the 99^{th} percentile. In the case of the 99^{th} percentile, only Lnkm and Porc_vurba present a significant effect, while the rest of the parameters are no longer significant at the 5% level, including the model intercept. The lack of significance may be explained by the wider confidence intervals at a 5% level of significance observed in Figure 2 for the 99^{th} percentile.

Second, Table 4 and Figure 2 also show that the magnitude of the marginal effects of variables with significant parameters in the models differs depending on the level of the estimated quantile. Specifically, the marginal effect of Lnkm increases as the level of the estimated quantile increases (being equal to 597.6 and 1180.2 for the 50^{th} and 99^{th} percentiles, respectively). The same pattern, albeit less pronounced, is observed for the marginal effect of Porc_nocturn, which increases as the level of the estimated quantile increases (being equal to 5.41 and 37.49 for the 50^{th} and 95^{th} percentiles, respectively). In the case of Porc_vurba, the marginal effect is always negative, but in absolute terms it is increasing with the level of the estimated quantile (being equal to -9.19 and -87.12 for the 50^{th} and 99^{th} percentiles, respectively). Finally, the marginal effect of gender is always positive and is increasing with the level of the estimated quantile (being equal to 206.76 and 1070.06 for the 50^{th} and 97.5^{th}, respectively).

It is interesting to compare the results of the quantile regression for the 75^{th} and 95^{th} percentiles. Thus, the model intercept is quite similar in both models. A comparison of the marginal effect of Lnkm shows that a one-unit increase in Lnkm (equivalent to multiplying Km by 2.718), increases 892.80 km the 75^{th} percentile of the number of kilometers driven at speeds above the posted limit, while the 95^{th} percentile increases 1094.57 km, ceteris paribus. In the case of Porc_vurba, increasing the percentage of kilometers driven in urban areas by one percentage unit reduces 22.26 km the 75^{th} percentile of the number of kilometers driven at speeds above the posted limit, and 53.44 km the 95^{th} percentile, ceteris paribus. On the other hand, being a man increases 377.94 km the 75^{th} percentile of the number of kilometers driven at speeds above the posted limit and 755.87 km the 95^{th} percentile, ceteris paribus. Finally, increasing the percentage of kilometers driven at night by one percentage unit increases 6.71 km the 75^{th} percentile of the number of kilometers driven at speeds above the limit and 37.49 km the 95^{th} percentile, ceteris paribus.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2019 doi:10.20944/preprints201906.0072.v1

Peer-reviewed version available at *Risks* **2019**, *7*, 80; doi:10.3390/risks7030080

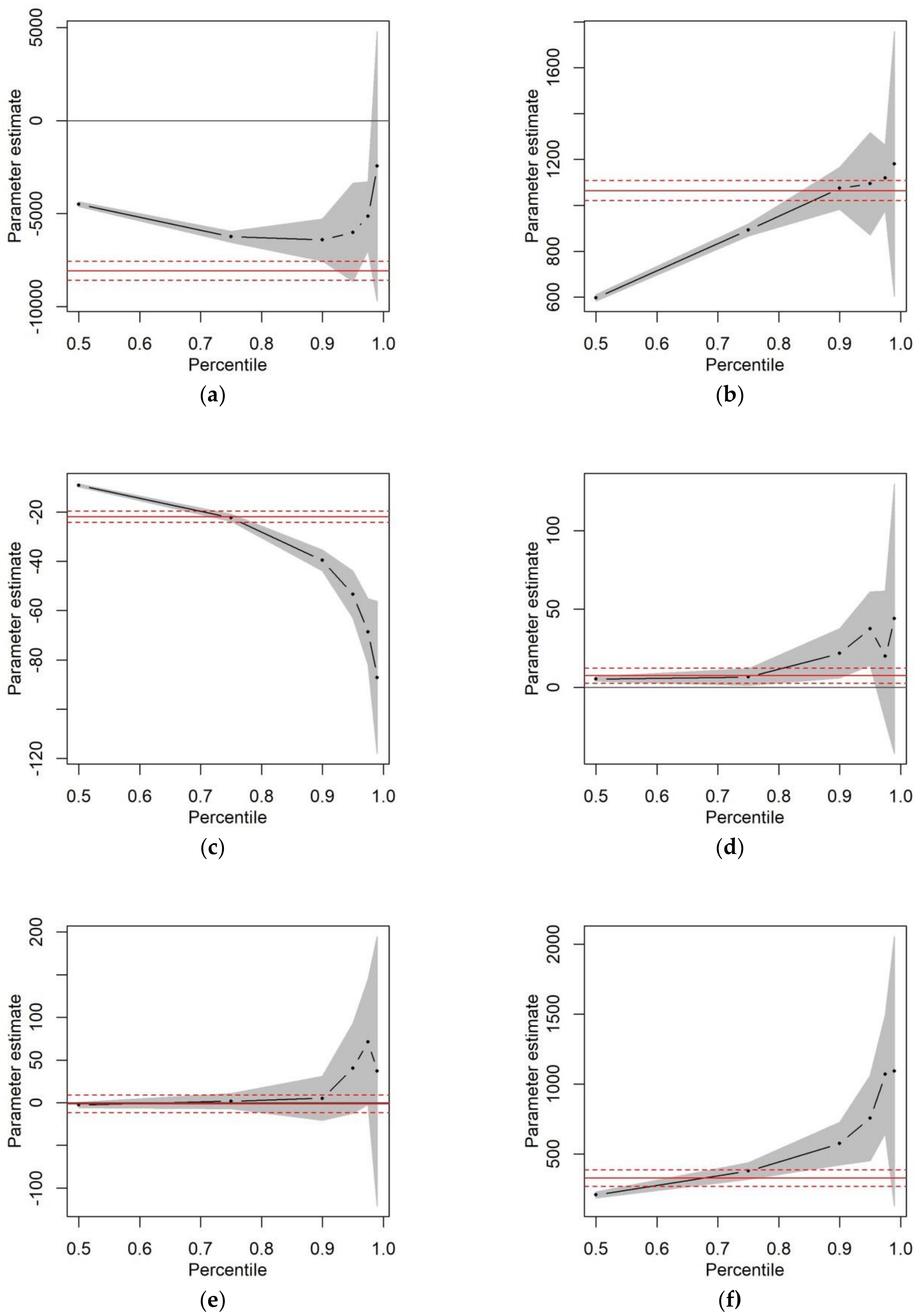

Figure 2. Parameter estimates at different levels of the quantile. Confidence intervals at a 5% level of significance. The horizontal red line represents the corresponding parameter estimate in a classical linear regression model. (a) Intercept; (b) lnkm; (c) porc_vurba; (d) porc_nocturn; (e) age; (f) gender.

Finally, Table 5 illustrates how the model can be implemented for predictive purposes. Let us consider three drivers with different characteristics, each of whom has driven exactly 600 km above the posted speed limit. Compared to the general population, and without conditioning on specific

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2019 doi:10.20944/preprints201906.0072.v1

Peer-reviewed version available at *Risks* **2019**, *7*, 80; doi:10.3390/risks7030080

characteristics, these three drivers present a distance driven at excess speeds below the median (689.20 km) and, as such, can be considered relatively safe drivers. However, the key is to calculate the percentile risk level of the response variable given the specific characteristics of each driver. Indeed, it seems obvious that a distance of 600 km driven above the posted speed limit does not denote the same level of risk for an urban driver (who probably does a lot of driving in congested areas), as it does for a driver who drives largely outside the city limits. Most notably, the risk depends on the total distance driven. If we use the grid of different percentiles (Table 4) to make our predictions, it can be seen that for a distance of 600 km driven above the speed limit, driver 1 lies at the 50^{th} percentile, indicative of median risk. In contrast, driver 2 lies at the 75^{th} percentile and, so, has a higher risk score when taking his driving characteristics into account. And, finally, driver 3 lies at the 90^{th} percentile, indicative of a very high risk.

Table 5. Estimates of the conditional percentiles for drivers with different characteristics, each of whom has driven 600 km above the posted speed limit.

	Driver 1	Driver 2	Driver 3
Km	12,000	8,000	5,500
Porc_vurba	80	75	80
Porc_noctur	14	11	10.5
Age	25	25	25
Gender	1	1	1
Estimated conditional percentile [1]	50^{th}	75^{th}	90^{th}

[1] The estimated conditional percentile is found by locating the quantile level that produces a response equal to 600 km, given the exogenous characteristics (total kilometers driven, percent urban driving, percent nighttime driving, age and gender) in the three example columns.

5. Conclusions

We have shown that the distribution of the distance driven above the posted speed limit is not homogeneous with respect to certain driver characteristics. As such, quantile regression is an interesting tool for analyzing risk when telematics information is available. On the assumption that quantiles of distance driven above the speed limit represent a valuable risk measure, our model allows us to identify the factors associated with higher quantile values and, therefore, with risky drivers. This information is valuable in terms of providing preventive early warnings.

We also find that the impact of each additional kilometer driven is much greater in higher quantiles than in lower quantiles. Note that we specify a log-linear relationship between total distance driven and distance driven above the posted speed limits, which means there is a decreasing marginal effect on the latter as total distance increases.

One limitation of our analysis is that the degree to which drivers exceeded the posted limit was not recorded by the telematics equipment; thus, we are unable to examine the magnitude of the speed violation.

We believe that UBI will soon develop into a scheme that can improve aspects of both service and protection in the sector. As insurance services are reinvented, risk scores and the identification of potential niches of drivers with risky patterns provide a new way of keeping drivers better informed and of promoting safe driving. Models such as those presented in this paper should enable insurers to design predictive models of driver risk and fix personalized indicators. In the application presented here, it could be argued that excess speed is the only feature a driver can modify, given that all other factors, including age, gender, total distance driven, and percentages of nighttime and urban driving, are dictated by external circumstances such as distance from home to work place, and by personal or professional obligations. This means the quantile regression model would predict the total distance driven above the posted speed limit percentile, given that particular set of external circumstances and, thus, it would allow the percentile risk score of the driver to be calculated by controlling for those circumstances and not for the whole population of drivers. Estimating a

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2019 doi:10.20944/preprints201906.0072.v1

Peer-reviewed version available at *Risks* **2019**, *7*, 80; doi:10.3390/risks7030080

driver's rank with regard to distance driven above the posted speed limit is personalized information that should constitute interesting feedback for policy holders. Indeed, safety measures and even telematics-based insurance should segment the population of drivers accordingly. Given that speed is the primary cause of severe accidents, these results should translate into lower insurance premiums for those who present a lower risk. In other words, if quantile-based behavior is considered rather than mathematical expectations of accident severity, the calculation of the premium to be paid should be improved. However, we leave questions as to how this rank might be converted into an insurance price and how information of a driver's behavior might impact careful driving for further research.

Author Contributions: Conceptualization, M. Alcañiz and L. Bermúdez; methodology, M. Guillen and A.M. Pérez-Marín.; software, Ana M. Pérez-Marín and M. Alcañiz; validation, M. Alcañiz; formal analysis, Ana M. Pérez-Marín; investigation, M. Guillen; resources, M. Guillen; data curation, L. Bermúdez; writing—original draft preparation, Ana M. Pérez-Marín and L. Bermúdez; writing—review and editing, L. Bermúdez; visualization, Pérez-Marín; supervision, M. Guillen; project administration, M. Guillen; funding acquisition, M. Guillen.

Funding: Support from the Spanish Ministry and ERDF grant ECO2016-76203-C2-2-P is gratefully acknowledged. MG gratefully acknowledges financial support from ICREA under the ICREA Academia programme.

Conflicts of Interest: The authors declare no conflict of interest.

References

(Ayuso et al., 2014) Ayuso, Mercedes, Guillen, Montserrat, and Pérez-Marín, Ana Maria. 2014. Time and distance to first accident and driving patterns of young drivers with pay-as-you-drive insurance. *Accident Analysis and Prevention* 73: 125–31. DOI: https://doi.org/10.1016/j.aap.2014.08.017.

(Ayuso et al., 2016a) Ayuso, Mercedes, Guillen, Montserrat, and Pérez-Marín, Ana Maria. 2016a. Telematics and gender discrimination: some usage-based evidence on whether men's risk of accident differs from women's. *Risks* 4:2: 10. DOI: https://doi.org/10.3390/risks4020010.

(Ayuso et al., 2016b) Ayuso, Mercedes, Guillen, Montserrat, and Pérez-Marín, Ana Maria. 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 Emerging Technologies* 68: 160–7. DOI: https://doi.org/10.1016/j.trc.2016.04.004.

(Boucher et al., 2013) Boucher, Jean-Philippe, Pérez-Marín, Ana Maria, and Santolino, Miguel. 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: 135-54.

(Dissanayake and Lu, 2002) Dissanayake, Susanda, and Lu, Jian John. 2002. Factors influential in making an injury severity difference to older drivers involved in fixed object-passenger car crashes. *Accident Analysis and Prevention* 34: 5: 609–18. DOI: https://doi.org/10.1016/S0001-4575(01)00060-4.

(Gao et al., 2019) Gao, Guangyuan, Meng, Shengwang, and Wüthrich, Mario V. 2019. Claims frequency modeling using telematics car driving data. *Scandinavian Actuarial Journal* 2019: 2: 143-62. DOI: https://doi.org/10.1080/03461238.2018.1523068.

(Gao and Wüthrich, 2019) Gao, Guangyuan, and Wüthrich, Mario V. 2019. Convolutional neural network classification of telematics car driving data. *Risks* 7: 1: 6. DOI: https://doi.org/10.3390/risks7010006.

(Guillen et al., 2019) Guillen, Montserrat, Nielsen, Jens Perch, Ayuso, Mercedes, and Pérez-Marín, Ana Maria. 2019. The use of telematics devices to improve automobile insurance rates. *Risk Analysis*, 39: 3: 662-72. DOI: https://doi.org/10.1111/risa.13172.

(Hewson, 2008) Hewson, Paul James, 2008. Quantile regression provides a fuller analysis of speed data. *Accident Analysis and Prevention* 40: 502–10. DOI: 10.1016/j.aap.2007.08.007.

(Jun et al., 2007) Jun, Jungwook, Ogle, Jennifer, and Guensler, Randall. 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: 246–55.

(Jun et al., 2011) Jun, Jungwook, Guensler, Randall, and Ogle, Jennifer. 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: 569–78. DOI: https://doi.org/10.1016/j.trc.2010.09.005.

(Koenker and Hallock, 2001) Koenker, Roger, and Hallock, Kevin. 2001. Quantile regression. *Journal of Economic Perspectives* 15: 143–56. DOI: 10.1257/jep.15.4.143.

(Koenker et al., 2018) Koenker, Roger, Portnoy, Stephen, Ng, Pin Tian, Zeisleis, Achim, Grosjean, Philip, and Ripley, Brian D. 2018. Package 'quantreg'. R Package Version 5.38, https://cran.r-project.org/web/packages/quantreg/quantreg.pdf.

(Ossiander and Cummings, 2002) Ossiander, Eric M., and Cummings, Peter. 2002. Freeway speed limits and traffic fatalities in Washington State. *Accident Analysis and Prevention* 34: 13–8. DOI: https://doi.org/10.1016/S0001-4575(00)00098-1.

(Paefgen et al., 2014) Paefgen, Johannes, Staake, Thorsten, and Fleisch, Elgar. 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. DOI: https://doi.org/10.1016/j.tra.2013.11.010.

(Pérez-Marín et al., 2019) Pérez-Marín, Ana Maria, Ayuso, Mercedes, and Guillen, Montserrat. 2019. Do young insured drivers slow down after suffering an accident?. *Transportation Research Part F: Traffic Psychology and Behaviour* 62: 690-99. DOI: https://doi.org/10.1016/j.trf.2019.02.021.

(Pérez-Marín and Guillen, 2019) Pérez-Marín, Ana Maria, and Guillen, Montserrat. 2019. Semi-autonomous vehicles: Usage-based data evidences of what could be expected from eliminating speed limit violations. Accident Analysis and Prevention 123: 99–106. DOI: https://doi.org/10.1016/j.aap.2018.11.005.

(Pilkington and Kinra, 2005) Pilkington, Paul, and Kinra, Sanjay. 2005. Effectiveness of speed cameras in preventing road traffic collisions and related casualties: systematic review. *BMJ* 330: 7487: 331–334. DOI: https://doi.org/10.1136/bmj.38324.646574.AE.

(Taylor et al., 2002) Taylor, M., Baruya, A., and Kennedy, J. 2002. The Relationship between speed and accidents on rural single-carriageway roads. *TRL Report TRL511. TRL*.

(Vernon et al., 2004) Vernon, Donald, Cook, Lawrence J., Peterson, Katherine J., and Dean, J. Michael. 2004. Effect of the repeal of the national maximum speed limit law on occurrence of crashes, injury crashes, and fatal crashes on Utah highways. *Accident Analysis and Prevention* 36: 223–9. DOI: https://doi.org/10.1016/S0001-4575(02)00151-3.

(Wilson et al., 2006) Wilson Cecilia, Willis Charlene, Hendrikz Joan K., and Bellamy Nicholas. 2006. Speed enforcement detection devices for preventing road traffic injuries. *Cochrane Database of Systematic Reviews Issue* 2. Art. No.: CD004607, doi:10.1002/14651858.CD004607.pub2.

(Yu and Abdel-Aty, 2014) Yu, Rongjie, and Abdel-Aty, Mohamed. 2014. Using hierarchical Bayesian binary probit models to analyze crash injury severity on high speed facilities with real-time traffic data. *Accident Analysis and Prevention* 62: 161–7. DOI: https://doi.org/10.1016/j.aap.2013.08.009.

(Yu et al., 2003) Yu, Keming, Lu, Zudi, and Stander, Julian. 2003. Quantile regression: applications and current research areas. *Journal of the Royal Statistical Society D* 52: 331–50.