*Article*

# Weather Conditions and Telematics Panel Data in Monthly Motor Insurance Claim Frequency Models

**Jan Reig Torra, Montserrat Guillen \*, Ana M. Pérez-Marín, Lorena Rey Gámez and Giselle Aguer**

Department of Econometrics, Statistics and Applied Economics, RISKcenter-IREA, Universitat de Barcelona, 08034 Barcelona, Spain; janreig@mit.edu (J.R.T.); amperez@ub.edu (A.M.P.-M.)
\* Correspondence: mguillen@ub.edu; Tel.: +34-934037039

**Abstract:** Risk analysis in motor insurance aims to identify factors that increase the frequency of accidents. Telematics data is used to measure behavioural information of drivers. Contextual variables include temperature, rain, wind and traffic conditions that are external to the driver, but may also influence the probability of having an accident, as well as vehicle and personal characteristics. This paper uses a monthly panel data structure and the Poisson model to predict the expected frequency of claims over time. Some meteorological information is included. Two types of claims are considered separately: only those related to at-fault third-party liability accidents, and all types of claims including assistance on the road. A sample of drivers in Spain in 2018–2019 is analysed with information on claiming frequency per month. Drivers were observed for seven months. Our analysis is novel because monthly summaries of telematics information are combined with weather data in a panel structure, revealing that external factors affect the expected claims frequencies. Reckless speeding behaviours and intense urban circulation increase the risk of an accident, which also increases with windy conditions.

**Keywords:** motor insurance; predictive models; telematics data; contextual data; at-fault claims

## 1. Introduction

### 1.1. Context and Motivation

Frequency models in motor insurance can identify factors that are associated with an increase in the probability of accident occurrence and can help identify which features should be taken into consideration when calculating the pure premium of a motor insurance policy. Combined with severity models, frequency models are fundamental for motor insurance pricing and, as such, regulators scrutinise how these are constructed internally by insurers in order to guarantee that they comply with legal standards and are transparent.

New technologies based on on-board devices, smartphones or the introduction of sensors in vehicles have expanded the possibilities of collecting information on vehicles in motion and, therefore, of considering new features in the design of claim frequency models. These new sources of data have led to the development of what is called "motor insurance telematics". Telematics data have proven to be very effective in identifying the behavioural traits of the insured which, combined with traditional rating factors such as age or place of residence, can be used to assess the claims propensity of a certain driver. All in all, it is now widely accepted that telematics data can be introduced in claim frequency models to increase their predictive performance compared to models based exclusively on classical non-telematic rating factors such as age, vehicle power, age of the policyholder's driver's licence, vehicle age, and so on.

The large amount of information that we can expect to collect in a telematics landscape allows statisticians and data analysts to play a decisive role in risk analysis. However, transforming a vast amount of available data into relevant information remains a bigger challenge than anticipated a few years ago when the interest in telematics information intensified enormously.

Our aim here is to combine telematics information from vehicles in motion with external contextual information from meteorological databases. It is widely accepted that driving conditions may affect the way a driver operates their vehicle and that storms, heavy rain, wind and fog escalate the need to drive with maximum alertness. However, the combination of traditional rating factors, telematics information and weather conditions has not been analysed when modelling motor insurance claim frequencies. Usually, it is assumed that the place of residence indirectly captures the external atmospheric features that a driver usually encounters, because some places experience more rain than others or are more exposed to dangerous weather phenomena. We analyse a monthly panel of telematics motor insurance records together with information on wind, sunshine, and average temperature.

*1.2. Background*

Pay-as-you-drive (PAYD) schemes are the simplest form of motor insurance telematics pricing, where premiums basically increase with distance driven. Insurers base their predictive modelling on records of distance driven during a certain period of time, usually one year, together with traditional rating factors and observed claim frequency. It has been proven (see Cheng et al. 2022) that PAYD benefits the policyholder by reducing their premium payments and increasing the total utility derived from vehicle usage. Moreover, it also has positive effects for insurers, since it has been shown that early adopters experience a significant increase in their market share, as they are able to attract low-risk drivers from other insurers (see Che et al. 2021). As already mentioned, the most primitive telematics variable is the distance travelled, and its relationship with accident risk has been intensively investigated in the literature. Boucher et al. (2013) analysed real data and found that this relationship is not linear, as driving twice as much as other drivers with the same characteristics results in the expected number of claims being less than twice the number of accidents. This non-linearity was also observed in further studies (see Boucher et al. 2017; Guillen et al. 2019, among others). The so-called learning effect is one possible explanation, as those who drive more become more skilful at driving. Those who drive more kilometres are also probably using safer roads. More recently, Boucher and Turcotte (2020) analysed panel telematics count data using Generalised Additive Models (GAMs) and found that the distance–claims frequency relationship seems to be linear, and the apparent non-linearity was caused by a residual heterogeneity captured by GAMs. Another issue that has recently been investigated in the literature is the existence of a cut-off value of distance driven, below which drivers who have a traditional policy should switch to a PAYD policy (Cheng et al. 2022). The impact of non-linearities and interactions may be better captured with, for example, neural nets or the combined GLM/NN methods that have been proposed in the literature (see Blier-Wong et al. 2020; Owens et al. 2022). However, they use cross-sectional information on insurance claims, and no studies have used panel data.

The literature presents evidence of the association between other telematics variables and the risk of having an accident. There is evidence of a better assessment of driving behaviour when detailed telematics information is used compared with traditional risk factors (see, for example, So et al. 2021). Specifically, speed, road type, and the time of day when the vehicle is used seem to have a strong relationship with the number of claims, and alternative models have been proposed in the literature to describe the joint dynamics of telematics data and claim occurrence (Ayuso et al. 2014; Pérez-Marín and Guillen 2019; Pérez-Marín et al. 2019; Guillen et al. 2019; Corradin et al. 2022, among others). Moreover, such detailed telematics information can be used to identify periods of increased risk by classifying trips that occur immediately before a claim (see Williams et al. 2022). Eling and Kraft (2020) reviewed over fifty papers published in the last two decades on usage-based insurance, with a summary of the methods, the selected telematics variables, and the key findings. In light of Eling and Kraft's survey, it is clear that telematics infor-

mation is no longer in its infancy and is a useful tool for enhancing the performance of classical pricing methods.

However, managing telematics information for motor insurance is difficult due to its volume and complexity, requiring technical skills from IT departments and new perspectives from actuarial teams. The development of big data in the insurance sector requires information to be summarized and key risk factors identified for pricing purposes. Additionally, data cleaning and pre-processing also become more difficult to perform with telematics information (see Gao et al. 2022). In this context, Barry and Charpentier (2020) assess the impact of big data technologies on insurance ratemaking, with a special focus on motor policies. These authors also consider that the big volume of data and the personalization compromises behind usage-based insurance shakes the homogeneity hypothesis behind pooling. It is clear that telematics information enhances the capacity to differentiate or discriminate among the insured. Recently, Frees and Huang (2021) analysed the social and economic principles that can be used to assess the appropriateness of discrimination nowadays in the insurance sector.

In this context, different approaches have been proposed to deal with the huge amount of telematics data. On the one hand, new variable selection algorithms have been developed to deal with a large number of covariates in the context of motor insurance (see Chan et al. 2022). Additionally, research has investigated exactly how much information should be collected. Duval et al. (2022) found that telematics data become redundant after about 3 months or 4000 km of records, at least for risk classification. Another approach that somehow simplifies this task is the consideration of near-miss events in the driving context (an incident that might have resulted in an accident), as they can help to reduce the amount of information needed for risk classification. Guillen et al. (2020) identified the relationship between telematics information (speed, road type, and time of the day) and three types of near-miss events (harsh acceleration, braking, and cornering). Their results are relevant for claim risk quantification based on dynamic monitoring of near-miss events, but the way that near-miss events and tariffs should be connected is still an open debate. Finally, another approach that can help to summarize telematics information consists of calculating a driving behaviour score for each finished trip. In this regard, recently Meng et al. (2022) used a supervised driving risk-scoring neural network model to calculate a risk score. The authors showed that incorporating risk scores significantly improves the prediction performance of classical Poisson regression models for the frequency of claims.

Beyond behavioural information captured in telematics features, evidence has recently shown that contextual information is also potentially relevant when assessing the risk of suffering an accident, specifically in bad weather. Although this type of information is missing from most of the papers on usage-based insurance, many authors admit that traffic congestion and weather conditions, in addition to the time of day and type of road, influence the frequency and severity of claims. Qiu and Nixon (2008) found that crash rates increase during precipitation; specifically, they showed that snow has a greater effect than rain on crash occurrence (snow increases the crash rate by 84% and the injury rate by 75%). Mornet et al. (2015) used information from historical wind speed data in France to create an index to assess the economic impact of storms on insurance portfolios. Malin et al. (2019) investigated the relative accident risk of different road weather conditions as well as combinations of conditions. They found that relative accident risks were the highest for icy rain and slippery road conditions. Liang et al. (2021) found there to be a non-linear and lag relationship between local temperatures and traffic accident injuries; specifically, both low temperatures and high temperatures increase the incidence of traffic accident injuries. In addition, different temperatures have different effects on accident injuries for different age and gender groups. More recently, Gao and Shi (2022) also investigated the impact of hailstorms on hail damage claims using data from a U.S. insurer. The authors used information provided by hail radar maps and other spatially-varying weather features to help insurers anticipate and manage claims more efficiently.

To our knowledge, there has been no previous attempt to combine telematics and weather condition information in panel data for frequency models.

### 1.3. Objectives

We are interested in answering two research questions using telematics data. First, we aim to find a method for identifying meteorological factors that should be considered when modelling claim frequency. Second, by analysing a panel structure in our monthly data, we aim to confirm an autoregressive pattern that demonstrates whether a previous month's claiming behaviour and telematics information have an influence on the expected current number of claims.

In what follows, a claim is understood as an event not necessarily associated with an accident that requires the assistance of the insurer, such as technical assistance due to an incident (ignition or flat tyre, for example). We also aim to predict the frequency of at-fault accidents. Therefore, we compare the results of predictive models: (a) for the number of claims in general, and (b) for the number of at-fault third-party liability claims. Furthermore, we aim to identify the impact of a selection of weather conditions on the expected claim frequency for each of these two responses.

A real case study using telematics data is used to show the implementation in a panel of insured drivers observed over time on a monthly basis.

### 1.4. Data and Structure

The original database is composed of telematics data on insured drivers and contains a series of 249,408 observations. Each policy corresponds to an insured driver and information on different variables is obtained each month. 12 months of effective driving are observed for each insured driver, although these months may not necessarily be consecutive because some drivers use their cars quite irregularly. Only months when vehicles were used at least once a week are considered. This could bias the results because drivers who only drive occasionally may be underrepresented. The total number of observed policyholders is 20,784 and corresponds to a sample of drivers in Spain between 2018 and 2019. Working on monthly data with such a large number of policyholders adds an innovative component to the analysis and allows for complex modelling, in particular panel data models. The policyholders received a telematics on-board box to collect data, but the insurance product started operating under the pay-as-you-drive scheme after four months. Therefore, we consider only months 6 to 12. Month 5 was also discarded because we observed a higher claim frequency of claims in that particular month and we strongly believe that policyholders may have reported claims which actually occurred in month 4, in order to be covered by the new policy. In fact, the embargo period of four months may have created a new form of fraud. We also note that policies start at different points in time. Therefore, even if month number 6 may correspond to June for some insured drivers who started their contracts in January, this could also correspond to August for drivers who signed their contracts in March. We would like to remark that many papers use summaries of telematic information but do not analyse longitudinal information of the same policy, instead combining weekly telematics information with historical records.

The data that we manage have two levels of aggregation, since an insured driver has as many rows as months driven. Furthermore, the panel data is balanced, since for each insured driver we have the same number of months observed. The original database has more than 60 variables that are grouped into 3 categories:

-   The first category contains the variables related to the basic features of the insured, such as gender, age, vehicle power, and driver's licence age. These variables do not vary over time, so they are implicitly included in the panel models as individual effects. In this category, we also consider the location where the insured driver has mostly driven each month, which is generally constant over time because it is usually their place of residence. Specifically, we have an indicator of the province where the driver has driven the majority of the monthly total distance.

- The second category includes time-varying variables related to the drivers' behaviour and weather information. Weather information contains data on sunshine hours, average monthly temperatures, and wind. The drivers' information includes features such as maximum and average speed and distance travelled obtained from daily information. All variables are disaggregated by road type: highway, national road, regional road, and urban road.
- Finally, the third category includes claim frequency information.

One of the most innovative highlights of our analysis is the introduction of weather data. Three new variables are initially introduced in the database: average monthly temperature, average number of hours of sunshine per day, and average wind speed. Unfortunately, we do not have information on rain. They are all linked to the province and month where each insured driver has driven. The data for these three variables were taken from the Agencia Estatal de Meteorología (2022) in Spain. The list of variables finally considered in our analysis is shown in Table 1. We would like to remark that a selection of variables was undertaken previous to the current choice. As in similar studies, total distance driven is one of the main telematics features (Pitarque and Guillen 2022; Huang and Meng 2019). Speeding is also one of the telematics variables that has a tremendous effect on the expected frequency of accidents (Ma et al. 2018). We wanted to avoid multicollinearity in the models, so we could not introduce distance driven by type of road.

**Table 1.** Description of the variables finally considered in the analysis. They are all measured from the policyholder's monthly information. Spanish data set, seven complete months observed in 2018 and 2019.

| Variable | Description |
| --- | --- |
| Telematics variables | |
| Tele_km_total_urban | Kilometres travelled on urban roads |
| Tele_speed_mean_urban | Mean speed on urban roads |
| Tele_speed_mean_national | Mean speed on national roads |
| Tele_speed_mean_highway | Mean speed on the highways |
| Tele_speed_max_urban | Maximum speed on urban roads |
| Tele_speed_max_national | Maximum speed on national roads |
| Tele_speed_max_highway | Maximum speed on the highways |
| Weather conditions | |
| Temperature | Average monthly temperature (degrees Celsius) |
| Sun | Average number of hours of sunshine per day during the month |
| Wind | Average monthly wind speed (km/h) |
| Response | |
| Claims | Number of claims (of any type) |
| At-fault claims | Number of at-fault third-party liability claims |

*1.5. Description of the Data*

In Figure 1 the average maximum speed is represented as a function of the month, depending on the type of road, and separating policies with claims from those without claims. We observe that the average maximum speed for drivers with claims is slightly higher than for those without claims. We also observe that for highways and national roads, the average maximum speed remains very stable over time. The average maximum is around 132 km/h on highways (where the maximum possible posted limit is 120 km/h) and 102 km/h on national roads (where the posted limit is 90 km/h). However, for urban roads, the maximum speed is clearly above the posted limit (more than 80 km/h when the posted limit is 50 km/h) and this increases in the last months of the observation period. Figure 2 presents the monthly average mean speed by type of road. At the beginning of the observation period the average values were around 76 km/h on highways, 55.5 km/h on national roads, and 19 km/h on urban roads, and they change over time (we observe a decrease for highways and an increase for national and urban roads).
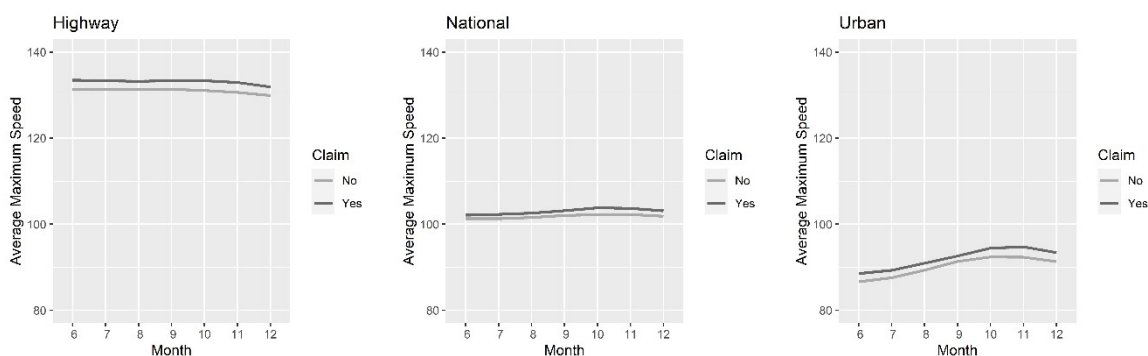
**Figure 1.** Plots of the average maximum speed per individual from month 6 to month 12 in terms of whether or not individuals reported claims at some point over the entire period. From left to right there is a graph for the maximum speed on highways, national roads, and urban roads, respectively.
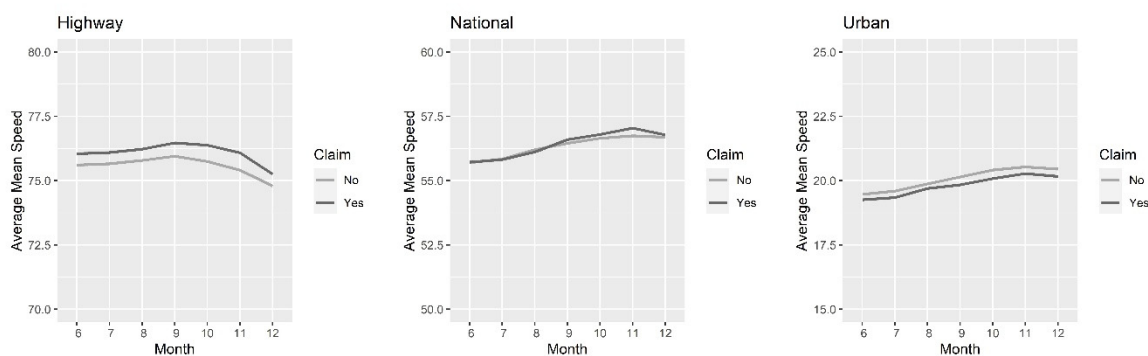


**Figure 2.** Plots of the average mean speed per individual from month 6 to month 12 in terms of whether or not individuals reported claims at some point over the entire period. From left to right there is a graph for the average mean speed on highways, national roads and urban roads, respectively.

Figure 3 presents the percentage monthly frequency of total claims and at-fault third-party liability claims together with the monthly averages of the temperature, wind speed, and sunshine daily hours. The frequency of claims decreases from month 6 to 11 (from 7.63% to 6.09% for all types of claims and from 0.41% to 0.24% for third-party liability claims), and then increases during the last month (reaching 6.75% for all claims and 0.35% for third-party liability claims). The temperature and sunshine daily hours, however, increase between months 6 and 10 and decrease at the end of the observation period, and wind speed slightly increases over time.

A full description of the data is presented in Table A1 in Appendix A. In Table 2, we also show a cross-table with the number of claims and at-fault accidents suffered by the insured in the sample.

Table 2 shows that 16,235 policyholders did not report any claim, while 2694 reported one claim, 475 two claims, and 1380 three or more claims. Note that those general claims may include all types of assistance. Only one policyholder reported three or more at-fault third-party liability claims. We can see that all the individuals who suffered an at-fault accident during the seven months also reported other claims. Furthermore, those insured who made fewer than three claims did not have any at-fault accidents. Therefore, those insured who made more claims are the ones who experienced at-fault accidents.
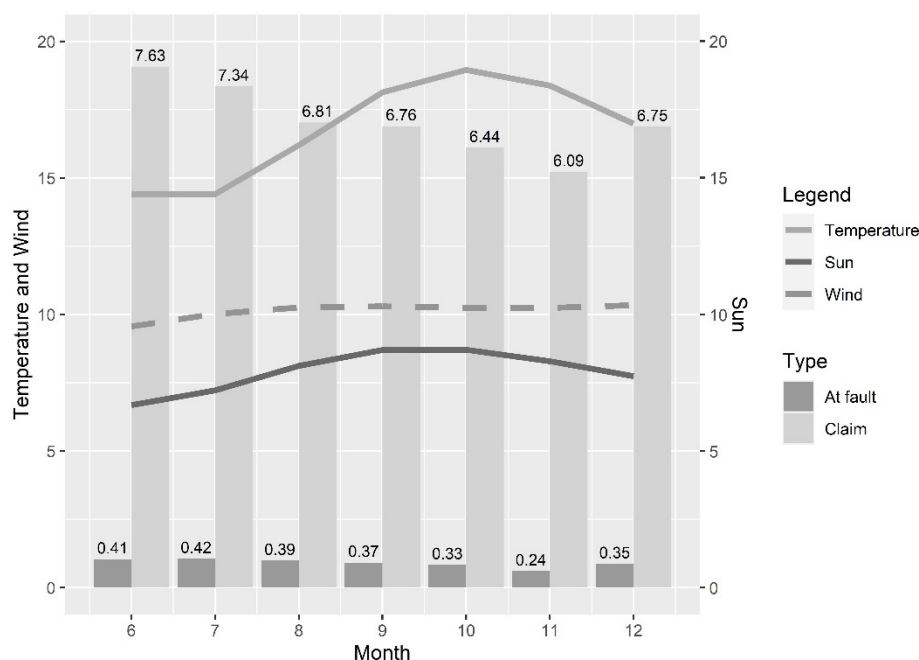
**Figure 3.** Percentage monthly frequency of claims (light grey) and percentage of at-fault third-party liability frequency of monthly claims from month 6 to month 12. The monthly average temperature in degrees Celsius (scale on the left axis) and monthly average wind speed in km/h (scale also on the left axis) are plotted, together with the average monthly sunshine daily hours (scale on the right axis).

**Table 2.** Cross-table with the total number of claims (rows) and at-fault third-party liability (TPL) claims (columns) per insured during the seven observed months in 2018 and 2019.

| All Claims | At-Fault TPL Claims | | | | Total |
|---|---|---|---|---|---|
| | **0** | **1** | **2** | **3 or More** | |
| 0 | 16,235 | 0 | 0 | 0 | 16,235 |
| 1 | 2694 | 0 | 0 | 0 | 2694 |
| 2 | 475 | 0 | 0 | 0 | 475 |
| 3 or more | 868 | 502 | 9 | 1 | 1380 |
| Total | 20,272 | 502 | 9 | 1 | 20,784 |

## 2. Methods

Using panel data for the study objectives is adequate because we have longitudinal information. Even if neural network or combined GLM/NN methods could provide newer insights into non-linearities, they do not provide the simplicity of linear models for an explanation of the effects.

We use the Poisson model for monthly panel data to predict the expected total number of claims and the expected total number of at-fault third-party liability policy claims. The model considers explanatory variables varying over time:

- Behavioural variables allow us to identify the most dangerous forms of conduct.
- Variables related to the weather conditions that change with the area and month where the vehicle has been driven allow us to see which environments are the most dangerous.

Individual non-time-changing factors constitute the individual effects.

### 2.1. Software and Package

The statistical analysis was carried out using R software. The estimation of the models with panel data was performed using the pglm() function of the pglm package released in July 2021 (see Croissant 2021). This package is part of the R software's current libraries and allows the estimation of generalised linear models with panel data.

### 2.2. Models Specification and Estimation

The model specification for the number of total claims, a response variable called $y_{it}$, for an individual $i$ in the driving month $t$, consists of a panel data counts model. It is assumed that conditional on the latent state variables $\lambda_{it}$, $y_{it}$ are mutually independent with Poisson distribution and with mean $\mu_{it} = \exp(\lambda_{it})$. This assumption means that we are estimating a random effects model. The corresponding probability density function of $y_{it}$ given $\lambda_{it}$, where $\lambda_{it}$ represents the logarithm of the conditional mean of $y_{it}$, from the Poisson distribution family, is:

$$f(y_{it}|\lambda_{it}) = \frac{\exp\{y_{it}\lambda_{it} - \exp(\lambda_{it})\}}{y_{it}!}, \; i = 1, \ldots, N \; t = 1, \ldots, T$$

where $N$ is the total number of individual policyholders and $T$ is the total number of months observed. The term $\lambda_t$ is autoregressive and is defined as:

$$\lambda_t = \kappa \, \lambda_{t-1} + X_t\beta + \epsilon_t \, , \tag{1}$$

where $\lambda_t = (\lambda_{it})$ is the $N \times 1$-dimensional vector of the state variables in period $t$, $\epsilon_t = (\epsilon_{it})$ is the $N \times 1$-dimensional vector of the error terms and $X_t$ is the $N \times K$ matrix where the $i$-th row $x_{it} = (x_{ikt})$ consists of the $K$ covariates observed by individual $i$ in month $t$.

In order to take into account possible unobserved and time-invariant heterogeneity we assume that the error term follows a Gaussian normal distribution of random effects of the form: $\epsilon_t = \tau + e_t$, where $\tau|X_t \sim N(0, \sigma_\tau^2)$ and $e_t|X_t \sim N(0, \sigma_e^2)$. The vector $\tau = (\tau_i)$ contains the specific individual effects that are not taken into account by the $X_t$ regressors and $e_t = (e_{it})$ is the vector of disturbance terms. It is assumed that $\tau$ and $i$ are mutually independent, conditional on $X_t$. Therefore, a random effects model is estimated where the individual effects $\tau_i$ are assumed to be random and distributed independently of the regressors.

As shown in (1), the model is dynamic since it depends directly on past values through $\lambda_{t-1}$, which represents the value of the state variables in the previous period. Thus, we consider that $y_t$ has a Poisson distribution with mean that is a function of the regressors $x_t$ and past claims information, $y_{t-1}$.

Additionally, a fixed-effect model is also considered and (1) contains N individual parameters, $a_i$, each one summarizing individual non-time-changing information. We use the so-called within estimator. Therefore, we use a Poisson regression model which is expressed in terms of the differences in each observation with respect to the average of all individuals in each time period.

All the covariates available in the database have been considered and introduced into the model linear predictor. Finally, features have been selected using a backwards variable selection process and by minimizing the Akaike information criterion.

### 3. Materials and Methods

The main results of the models for random effects and for fixed effects are shown in Table 3, separated into a response that counts all observed claims and one that counts only at-fault third-party liability claims. In Table 3, only the monthly wind speed average is included as a weather feature. The results for average monthly sunshine hours and monthly temperature are presented in Tables A2 and A3 in Appendix A. Additionally, extreme temperature, sunshine and wind have also been considered. Specifically, we have repeated the analysis with monthly maximum averages and the conclusions do not vary substantially (see Tables A4–A6 in Appendix A).

**Table 3.** Cross-table with the parameter estimates of random effects and fixed-effects Poisson panel data models, with *p*-values, for the total number of claims (left) and at-fault third-party liability (TPL) claims (right) per month and per insured based on the seven observed months in 2018 and 2019.

| Variable | All Claims | | | | Only At-Fault TPL Claims | | | |
|---|---|---|---|---|---|---|---|---|
| | Random Effects | | Fixed Effects | | Random Effects | | Fixed Effects | |
| | Parameter | *p*-Value | Parameter | *p*-Value | Parameter | *p*-Value | Parameter | *p*-Value |
| Intercept | −2.942 | <0.001 | - | - | −7.226 | <0.001 | - | - |
| lag(claims *) | −0.445 | <0.001 | −0.569 | <0.001 | −1.200 | 0.252 | −4.370 | <0.001 |
| Wind | 0.022 | 0.003 | 0.016 | 0.066 | 0.087 | 0.001 | 0.078 | 0.033 |
| Tele_km_total_urban ** | 0.123 | 0.012 | −0.212 | <0.001 | 0.789 | <0.001 | 0.071 | 0.764 |
| Tele_speed_mean_urban | −0.020 | <0.001 | −0.020 | <0.001 | −0.024 | 0.007 | −0.028 | 0.041 |
| Tele_speed_max_highway | 0.004 | <0.001 | −0.001 | 0.407 | 0.007 | 0.021 | −0.011 | 0.047 |
| $\sigma_\tau$ | 0.158 | <0.001 | - | - | 1.008 | 0.161 | - | - |
| AIC | 58,911.58 | | 26,833.79 | | 5779.10 | | 1420.46 | |

* we consider either all claims or only at-fault TPL claims, correspondingly; ** expressed in thousands of kilometres.

### 3.1. Poisson Panel Data Model with Random Effects for the Number of Claims

Table 3 shows a summary of the final models with information on the linear predictor coefficients of the most relevant variables for the model, with random effects for the total number of claims as well as their significance test. The AIC of the model is 58,911.58.

The coefficient of the lag of the previous period claim observation is significant, as well as the sigma coefficient, except when at-fault TPL claims are being modelled with random effects. The sigma coefficient corresponds to the standard deviation of the time-invariant unobserved heterogeneity $\sigma_\tau$, i.e., the variance of the random individual effects. The significance of this coefficient tells us that there are differences between policyholders in the portfolio, or that it captures effects omitted from the model or imperfectly measured.

The fact that the coefficient of the time lag is significant and negative indicates that if the number of claims in period $t − 1$ was high, then the number of claims in the current period $t$ is expected to be lower and, conversely, if no claim was observed in period $t − 1$, the likelihood of observing a claim in period t increases. This result is evidence of an autoregressive pattern and is to be expected. If in one month a driver has at least one incident on the road, in the following months they may drive less or behave more cautiously to avoid new incidents. Likewise, if no incident is reported, then drivers may become more confident in the subsequent month.

The behaviours that pose the highest risk of incidents are high maximum speeds on highways and driving in urban areas, such as travelling over the posted speed limit. The coefficient of average speed on urban roads is seen to be significant but negative, which indicates that the risk of an incident is higher when driving at a lower average speed. This can be attributed to driving on more congested roads, or to driving less safely. We also observe that the total distance travelled on urban roads increases the number of claims.

Regarding the variables related to meteorological conditions, we conclude that windy conditions impact the risk of reporting a claim even when considering constant individual features and behavioural information. However, no evidence can be found for the effects of average temperature and average daily sunshine hours. The positive coefficient of wind indicates that driving in areas with wind or during windy months increases claim risk.

The coefficients $\beta$ of each variable can be interpreted approximately as semi-elasticities, giving the proportional increase in $E(y_{it}|x_{it})$ associated with a unit increase in the regressor $x_{it}$. To give an example, the coefficient of the maximum highway speed equals 0.004 in the random effects models (see Table 3, first column). Therefore, a one-kilometre per hour increase in the maximum highway speed per month is associated with a 0.4% increase in the expected total number of claims, given that the other regressors remain constant.

### 3.2. Poisson Panel Data Model with Fixed Effects for the Number of Claims

Table 3 also shows the results for the Poisson regression model with fixed effects for the total number of claims. The AIC of the model with fixed effects is 26,833.79, which is lower than the AIC obtained for the random effects structure. In this case, the model is expressed in terms of the differences between each observation and the corresponding mean of all individuals for each time period.

Again, the coefficient of the lag of the number of claims in the previous period is significant and negative. Therefore, again, we conclude that the number of claims fluctuates over time. In this case, the average monthly wind speed is not associated with a higher risk of accident. However, again, we observe a negative and significant coefficient for the mean speed on urban roads, while the maximum speed on highways does not have a significant effect. Finally, in this case, the total distance travelled on urban roads decreases the number of claims, therefore there is a change in the sign of this coefficient with respect to the random effects approach, and it is surprising. A possible explanation is that in the fixed effects model for all claims we have probably captured the behaviour of the driver and their personal characteristics.

### 3.3. Poisson Panel Data Model with Random Effects for At-Fault Third-Party Liability Claims

Table 3 shows the results for the Poisson model with random effects for the at-fault third-party liability claims. The AIC is 5779.10. In this case, the lag of the number of at-fault accidents in the previous period does not have a significant effect. The reason could be that third-party liability claims are too infrequent to detect the autoregressive pattern. Again, we conclude that wind increases the accident risk. High maximum speed on highways is also associated with a higher risk of accident than when maximum speed is lower. Regarding the mean speed on urban roads, we note that it has a negative and significant coefficient. Therefore, driving at a lower mean speed on urban roads increases the risk of accident (this effect can be explained again by the higher risk of accidents when driving in more congested areas, where the mean speed is lower). The rest of the telematics variables were finally included in the model for comparability, but were not selected by the variable selection algorithm. Too many explanatory variables induce multicollinearity. Lastly, we conclude that the sigma coefficient (variance of the random individual effects) is not significant, therefore there are no differences between individuals, just heterogeneity.

### 3.4. Poisson Panel Data Model with Fixed Effects for At-Fault Third-Party Liability Claims

Table 3 shows the results of the Poisson regression model with fixed effects for the at-fault accidents. In this case, the AIC equals 1420.46, which is lower than the AIC obtained for the random effects structure. Note that models with the lowest AIC contain insignificant terms, but we wanted to have the same covariates in all models to compare the results. In that model, we observe that the lag of the number of at-fault accidents has a negative and significant parameter. Again, we observe that wind is associated with a higher risk of at-fault accidents. High values of the maximum speed on highways and high mean urban speed reduce the risk of at-fault accidents, which can only be explained by the lack of congestion and heavy traffic.

Finally, we would like to mention that we also performed the calculations with Month 5 included in the analysis. The results are shown in Appendix A (Tables A7–A9). The conclusions do not change. Moreover, we have also analysed the results when interactions between weather conditions and driver behaviour are included in the models. The results are presented in the Appendix A (Tables A10–A12). We found an interesting interaction between sunshine hours and telematics information when we modelled the total number of claims. Specifically, we observe that the total distance travelled on urban roads has a negative and significant parameter, but sunshine hours and its interaction with the total distance travelled on urban roads have a positive and significant parameter. This means that driving on urban roads in combination with sunshine conditions contributes to increasing the number of claims. On the other hand, the interaction between sunshine hours and the

mean speed on urban roads is negative; therefore, driving at a higher mean speed on urban roads in combination with sunshine conditions contributes to decreasing the number of claims. Finally, we also observe that the maximum speed on highways has a positive and significant parameter, but its interaction with sunshine hours is significant and negative. This means that, although sunshine conditions and high values of the maximum speed on highways increase the number of claims separately, this effect is, to some extent, mitigated when both conditions occur at the same time.

## 4. Discussion and Conclusions

We investigate the effect of telematics variables and contextual information of weather conditions on the risk of claims and at-fault accidents using real panel data based on monthly information. The telematics information on speeding, which together with total distance driven is probably the most relevant telematics variable, shows that the type of road where speeding takes place is significant. We distinguish between the mean and maximum speed on three different types of roads (urban, national, and highway). Two modelling approaches are used: fixed vs. random effects.

For the model for all types of claims, we reach similar conclusions about the effect of the covariates for both modelling approaches. With regard to weather conditions, only a higher average wind speed is associated with a higher risk of accidents (for the random effects model). Contrary to what was expected, we do not find a significant association between claims and average monthly temperature and average sunshine hours. We would have expected that winter time, with low temperatures in Spain, would be riskier than summer time, but it appears that only windy conditions were relevant. We would like to mention that working with average values over a month for meteorological conditions is somewhat coarse, and is a limitation of analysing monthly data.

We also find a negative association between the number of claims and the average speed in urban areas, which can be explained by the higher risk of collision for those who normally drive on congested urban roads. There are articles in the literature that found the same result. Specifically, Abdel-Aty and Radwan (2000) showed that heavy traffic volume, speeding, narrow lane width, larger number of lanes, urban roadway sections, narrow shoulder width and reduced median width increase the likelihood of accident involvement. Cabrera-Arnau et al. (2020) found that minor and serious accidents are more frequent in urban areas, whereas fatal accidents are more likely in rural areas. Indeed, when looking at data from different parts of the world, it is well-known that accidents with fatalities are more frequent in rural areas, rather than urban areas (Zwerling et al. 2005). We assume maximum speed to be an indicator of risky driving (especially when it is above the posted limit of the road). The maximum speed on highways increases the risk of claims. Finally, for both modelling approaches, we also observe that the number of claims shows an autoregressive fluctuation over time, as having a claim in the previous month decreases the risk of having a claim in the subsequent month.

Regarding the number of at-fault accidents, we reach similar conclusions for both modelling approaches. Firstly, we observe that high temperatures and sunshine do not seem to be associated with fewer accidents, but we confirm that the lower the wind speed average, the lower the expected frequency of at-fault accidents. We also observe that higher values for the total distance driven on urban roads increases the risk of at-fault accidents (for the random effects model), but the mean speed on urban roads has the opposite effect, as lower values are associated with a higher risk of at-fault accidents. A possible explanation is, again, related to situations in which the mean speed is lower due to congestion, which facilitates more accidents. Finally, for random and fixed effects we find contradictory results regarding the effect of average maximum speed on highways.

Our findings highlight the importance of considering detailed telematics information in combination with contextual information in order to predict the risk of at-fault accidents or the risk of all types of claims. Both sources of information have been considered in accident risk research, but there is a lack of studies in which they are considered simul-

taneously and with a panel data structure including detailed telematics information, in particular when considering interactions. Therefore, even if we make a contribution to the prediction of accident risk, which can be used to reconsider pricing usage-based insurance products and also for designing road safety plans, considering contextual factors such as the weather is necessary to understand telematics information.

## Appendix A

**Table A1.** Descriptive statistics: mean and standard deviation (SD) for 20,784 policyholders in seven observed months in 2018 and 2019.

| Mean | Month 6 | Month 7 | Month 8 | Month 9 | Month 10 | Month 11 | Month 12 | Total |
|---|---|---|---|---|---|---|---|---|
| Telematics variables | | | | | | | | |
| Tele_km_total_urban | 361.745 | 353.987 | 345.384 | 331.785 | 322.242 | 303.230 | 283.157 | 328.790 |
| Tele_speed_mean_urban | 19.422 | 19.540 | 19.833 | 20.077 | 20.340 | 20.476 | 20.384 | 20.010 |
| Tele_speed_max_highway | 131.781 | 131.704 | 131.712 | 131.793 | 131.557 | 131.121 | 130.341 | 131.430 |
| Weather conditions | | | | | | | | |
| Wind | 9.560 | 10.021 | 10.259 | 10.297 | 10.251 | 10.246 | 10.343 | 10.140 |
| Temperature | 14.404 | 14.407 | 16.202 | 18.138 | 18.959 | 18.386 | 16.986 | 16.783 |
| Sun | 6.682 | 7.219 | 8.118 | 8.702 | 8.706 | 8.289 | 7.740 | 7.922 |
| Responses | | | | | | | | |
| Claims | 0.076 | 0.073 | 0.068 | 0.068 | 0.064 | 0.061 | 0.068 | 0.068 |
| At-fault third-party liability | 0.004 | 0.004 | 0.004 | 0.004 | 0.003 | 0.002 | 0.003 | 0.004 |
| **S.D.** | **Month 6** | **Month 7** | **Month 8** | **Month 9** | **Month 10** | **Month 11** | **Month 12** | **Total** |
| Telematics variables | | | | | | | | |
| Tele_km_total_urban | 310.455 | 297.183 | 293.006 | 275.432 | 265.530 | 245.056 | 230.455 | 230.455 |
| Tele_speed_mean_urban | 5.524 | 5.598 | 5.715 | 5.764 | 5.928 | 6.056 | 6.075 | 6.075 |
| Tele_speed_max_highway | 15.583 | 15.527 | 15.618 | 16.002 | 16.131 | 16.737 | 16.934 | 16.934 |
| Weather conditions | | | | | | | | |
| Wind | 1.802 | 1.877 | 1.734 | 1.666 | 1.657 | 1.739 | 1.864 | 1.864 |
| Temperature | 4.936 | 5.311 | 5.827 | 5.894 | 5.864 | 6.024 | 6.080 | 6.080 |
| Sun | 2.078 | 2.348 | 2.432 | 2.337 | 2.319 | 2.372 | 2.364 | 2.364 |
| Responses | | | | | | | | |
| Claims | 0.491 | 0.481 | 0.467 | 0.466 | 0.440 | 0.407 | 0.443 | 0.443 |
| At-fault third-party liability | 0.064 | 0.065 | 0.063 | 0.061 | 0.058 | 0.049 | 0.059 | 0.059 |

**Table A2.** Cross-table with the parameter estimates of random effects and fixed-effects Poisson panel data models, with *p*-values, for the total number of claims (left) and at-fault third-party liability (TPL) claims (right) per month and per insured based on the seven observed months in 2018 and 2019. The models consider the average number of sunshine hours per day as explanatory variable describing weather conditions.

| Variable | All Claims | | | | Only At-Fault TPL Claims | | | |
|---|---|---|---|---|---|---|---|---|
| | Random Effects | | Fixed Effects | | Random Effects | | Fixed Effects | |
| | Parameter | *p*-Value | Parameter | *p*-Value | Parameter | *p*-Value | Parameter | *p*-Value |
| Intercept | −2.688 | <0.001 | - | - | −6.049 | <0.001 | - | - |
| lag(claims *) | −0.445 | <0.001 | −0.569 | <0.001 | −1.177 | 0.261 | −4.357 | <0.001 |
| Sun | 0.001 | 0.979 | 0.004 | 0.438 | −0.025 | 0.209 | −0.003 | 0.887 |
| Tele_km_total_urban ** | 0.116 | 0.018 | −0.223 | <0.001 | 0.789 | <0.001 | 0.015 | 0.951 |
| Tele_speed_mean_urban | −0.020 | <0.001 | −0.020 | <0.001 | −0.022 | 0.014 | −0.028 | 0.043 |
| Tele_speed_max_highway | 0.003 | <0.001 | −0.001 | 0.328 | 0.006 | 0.048 | −0.011 | 0.045 |
| $\sigma_\tau$ | 0.158 | <0.001 | - | - | 1.000 | 0.159 | - | - |
| AIC | 58,920.18 | | 26,836.56 | | 5787.84 | | 1426.87 | |

\* we consider either all claims or only at-fault TPL claims, correspondingly; \*\* expressed in thousands of kilometres.

**Table A3.** Cross-table with the parameter estimates of random effects and fixed-effects Poisson panel data models, with *p*-values, for the total number of claims (left) and at-fault third-party liability (TPL) claims (right) per month and per insured based on the seven observed months in 2018 and 2019. The models consider the average monthly temperature as explanatory variable describing weather conditions.

| Variable | All Claims | | | | Only At-Fault TPL Claims | | | |
|---|---|---|---|---|---|---|---|---|
| | Random Effects | | Fixed Effects | | Random Effects | | Fixed Effects | |
| | Parameter | *p*-Value | Parameter | *p*-Value | Parameter | *p*-Value | Parameter | *p*-Value |
| Intercept | −2.680 | <0.001 | - | - | −6.022 | <0.001 | - | - |
| lag(claims *) | −0.444 | <0.001 | −0.570 | <0.001 | −1.177 | 0.261 | −4.357 | <0.001 |
| Temperature | −0.001 | 0.788 | 0.002 | 0.331 | −0.013 | 0.116 | −0.002 | 0.857 |
| Tele_km_total_urban ** | 0.116 | 0.017 | −0.224 | <0.001 | 0.797 | <0.001 | 0.016 | 0.946 |
| Tele_speed_mean_urban | −0.020 | <0.001 | −0.020 | <0.001 | −0.021 | 0.016 | −0.028 | 0.043 |
| Tele_speed_max_highway | 0.003 | <0.001 | −0.001 | 0.317 | 0.005 | 0.056 | −0.011 | 0.046 |
| $\sigma_\tau$ | 0.158 | <0.001 | - | - | 1.004 | 0.160 | - | - |
| AIC | 58,920.11 | | 26,836.22 | | 5786.94 | | 1426.86 | |

\* we consider either all claims or only at-fault TPL claims, correspondingly; \*\* expressed in thousands of kilometres.

**Table A4.** Cross-table with the parameter estimates of random effects and fixed-effects Poisson panel data models, with *p*-values. for the total number of claims (left) and at-fault third party liability (TPL) claims (right) per month and per insured based on the seven observed months in 2018 and 2019. Monthly maximum daily average wind speed is introduced as the weather-related regressor.

| Variable | ALL Claims | | | | Only at Fault TPL Claims | | | |
|---|---|---|---|---|---|---|---|---|
| | Random Effects | | Fixed Effects | | Random Effects | | Fixed Effects | |
| | Parameter | *p*-Value | Parameter | *p*-Value | Parameter | *p*-Value | Parameter | *p*-Value |
| Intercept | −2.839 | <0.001 | - | - | −6.920 | <0.001 | - | - |
| lag(claims *) | −0.444 | <0.001 | −0.569 | <0.001 | −1.187 | 0.257 | −4.358 | <0.001 |
| Wind max | 0.011 | 0.038 | 0.006 | 0.279 | 0.053 | 0.010 | 0.036 | 0.165 |
| Tele_km_total_urban ** | 0.120 | 0.014 | −0.218 | <0.001 | 0.788 | <0.001 | 0.039 | 0.870 |
| Tele_speed_mean_urban | −0.020 | <0.001 | −0.020 | <0.001 | −0.023 | 0.009 | −0.028 | 0.045 |
| Tele_speed_max_highway | 0.003 | <0.001 | −0.001 | 0.381 | 0.006 | 0.034 | −0.011 | 0.043 |
| $\sigma_\tau$ | 0.158 | <0.001 | - | - | 1.008 | 0.161 | - | - |
| AIC | 58,915.90 | | 26,836.00 | | 5782.95 | | 1424.98 | |

\* we consider either all claims or only at fault TPL claims, correspondingly; \*\* expressed in thousands of kilometres.

**Table A5.** Cross-table with the parameter estimates of random effects and fixed-effects Poisson panel data models, with *p*-values, for the total number of claims (left) and at-fault third party liability (TPL) claims (right) per month and per insured based on the seven observed months in 2018 and 2019. The models consider the monthly maximum daily average of sunshine hours as explanatory variable describing weather conditions.

| Variable | All Claims | | | | Only At Fault TPL Claims | | | |
|---|---|---|---|---|---|---|---|---|
| | Random Effects | | Fixed Effects | | Random Effects | | Fixed Effects | |
| | Parameter | *p*-Value | Parameter | *p*-Value | Parameter | *p*-Value | Parameter | *p*-Value |
| Intercept | −2.723 | <0.001 | - | - | −6.032 | <0.001 | - | - |
| lag(claims *) | −0.445 | <0.001 | −0.570 | <0.001 | −1.177 | 0.261 | −4.360 | <0.001 |
| Sun max | 0.005 | 0.391 | 0.009 | 0.088 | −0.022 | 0.284 | 0.009 | 0.724 |
| Tele_km_total_urban ** | 0.115 | 0.019 | −0.223 | <0.001 | 0.788 | <0.001 | 0.017 | 0.943 |
| Tele_speed_mean_urban | −0.020 | <0.001 | −0.020 | <0.001 | −0.022 | 0.014 | −0.029 | 0.037 |
| Tele_speed_max_highway | 0.003 | 0.001 | −0.001 | 0.295 | 0.006 | 0.047 | −0.011 | 0.040 |
| $\sigma_\tau$ | 0.158 | <0.001 | - | - | 1.001 | 0.160 | - | - |
| AIC | 58,919.45 | | 26,834.25 | | 5788.27 | | 1426.77 | |

\* we consider either all claims or only at fault TPL claims, correspondingly; ** expressed in thousands of kilometres.

**Table A6.** Cross-table with the parameter estimates of random effects and fixed-effects Poisson panel data models, with *p*-values, for the total number of claims (left) and at-fault third party liability (TPL) claims (right) per month and per insured based on the seven observed months in 2018 and 2019. The models consider the monthly maximum daily average temperature as explanatory variable describing weather conditions.

| Variable | All Claims | | | | Only at Fault TPL Claims | | | |
|---|---|---|---|---|---|---|---|---|
| | Random Effects | | Fixed Effects | | Random Effects | | Fixed Effects | |
| | Parameter | *p*-Value | Parameter | *p*-Value | Parameter | *p*-Value | Parameter | *p*-Value |
| Intercept | −2.702 | <0.001 | - | - | −5.981 | <0.001 | - | - |
| lag(claims *) | −0.445 | <0.001 | −0.570 | <0.001 | −1.176 | 0.261 | −4.357 | <0.001 |
| Temperature max | 0.001 | 0.581 | 0.004 | 0.053 | −0.014 | 0.084 | −0.001 | 0.881 |
| Tele_km_total_urban ** | 0.114 | 0.020 | −0.229 | <0.001 | 0.805 | <0.001 | 0.017 | 0.942 |
| Tele_speed_mean_urban | −0.020 | <0.001 | −0.020 | <0.001 | −0.021 | 0.016 | −0.028 | 0.042 |
| Tele_speed_max_highway | 0.003 | <0.001 | −0.001 | 0.273 | 0.006 | 0.053 | −0.011 | 0.046 |
| $\sigma_\tau$ | 0.158 | <0.001 | - | - | 1.007 | 0.161 | - | - |
| AIC | 58,919.88 | | 26,833.42 | | 5786.42 | | 1426.87 | |

\* we consider either all claims or only at fault TPL claims, correspondingly; ** expressed in thousands of kilometres.

**Table A7.** Cross-table with the parameter estimates of random effects and fixed-effects Poisson panel data models, with *p*-values. for the total number of claims (left) and at-fault third party liability (TPL) claims (right) per month and per insured based on the eight observed months in 2018 and 2019, i.e., month 5 is included.

| Variable | All Claims | | | | Only at Fault TPL Claims | | | |
|---|---|---|---|---|---|---|---|---|
| | Random Effects | | Fixed Effects | | Random Effects | | Fixed Effects | |
| | Parameter | *p*-Value | Parameter | *p*-Value | Parameter | *p*-Value | Parameter | *p*-Value |
| Intercept | −3.049 | <0.001 | - | - | −7.309 | <0.001 | - | - |
| lag(claims *) | −0.421 | <0.001 | −0.535 | <0.001 | −1.373 | 0.184 | −4.389 | <0.001 |
| Wind | 0.019 | 0.006 | 0.014 | 0.055 | 0.077 | 0.002 | 0.063 | 0.054 |
| Tele_km_total_urban ** | 0.064 | 0.145 | −0.275 | <0.001 | 0.727 | <0.001 | −0.042 | 0.842 |
| Tele_speed_mean_urban | 0.016 | <0.001 | −0.016 | <0.001 | −0.018 | 0.024 | −0.018 | 0.148 |
| Tele_speed_max_highway | 0.004 | <0.001 | <0.001 | 0.911 | 0.008 | 0.004 | −0.007 | 0.145 |
| $\sigma_\tau$ | 0.183 | <0.001 | - | - | 1.151 | 0.125 | - | - |
| AIC | 70,710.83 | | 34,887.78 | | 6893.57 | | 1886.79 | |

\* we consider either all claims or only at fault TPL claims, correspondingly; ** expressed in thousands of kilometres.

**Table A8.** Cross-table with the parameter estimates of random effects and fixed-effects Poisson panel data models, with *p*-values, for the total number of claims (left) and at-fault third party liability (TPL) claims (right) per month and per insured based on the eight observed months in 2018 and 2019, i.e., month 5 is included. The models consider the average number of sunshine hours per day as explanatory variable describing weather conditions.

| Variable | All Claims | | | | Only at Fault TPL Claims | | | |
|---|---|---|---|---|---|---|---|---|
| | Random Effects | | Fixed Effects | | Random Effects | | Fixed Effects | |
| | Parameter | *p*-Value | Parameter | *p*-Value | Parameter | *p*-Value | Parameter | *p*-Value |
| Intercept | −2.842 | <0.001 | - | - | −6.356 | <0.001 | - | - |
| lag(claims *) | −0.420 | <0.001 | −0.534 | <0.001 | −1.354 | 0.190 | −4.375 | <0.001 |
| Sun | 0.001 | 0.834 | 0.004 | 0.425 | −0.013 | 0.462 | 0.001 | 0.944 |
| Tele_km_total_urban ** | 0.059 | 0.177 | −0.283 | <0.001 | 0.731 | <0.001 | −0.081 | 0.701 |
| Tele_speed_mean_urban | −0.016 | <0.001 | −0.016 | <0.001 | −0.016 | 0.050 | −0.018 | 0.149 |
| Tele_speed_max_highway | 0.004 | <0.001 | >−0.001 | 0.965 | 0.007 | 0.011 | −0.007 | 0.120 |
| $\sigma_\tau$ | 0.183 | <0.001 | - | - | 1.139 | 0.123 | - | - |
| AIC | 70,718.47 | | 34,890.80 | | 6902.71 | | 1890.49 | |

* we consider either all claims or only at fault TPL claims, correspondingly; ** expressed in thousands of kilometres.

**Table A9.** Cross-table with the parameter estimates of random effects and fixed-effects Poisson panel data models, with *p*-values, for the total number of claims (left) and at-fault third party liability (TPL) claims (right) per month and per insured based on the eight observed months in 2018 and 2019, i.e., month 5 is included. The models consider the average monthly temperature as explanatory variable describing weather conditions.

| Variable | All Claims | | | | Only at Fault TPL Claims | | | |
|---|---|---|---|---|---|---|---|---|
| | Random Effects | | Fixed Effects | | Random Effects | | Fixed Effects | |
| | Parameter | *p*-Value | Parameter | *p*-Value | Parameter | *p*-Value | Parameter | *p*-Value |
| Intercept | −2.817 | <0.001 | - | - | −6.283 | <0.001 | - | - |
| lag(claims *) | −0.420 | <0.001 | −0.534 | <0.001 | −1.355 | 0.190 | −4.373 | <0.001 |
| Temperature | −0.002 | 0.377 | <0.001 | 0.982 | −0.010 | 0.163 | −0.005 | 0.545 |
| Tele_km_total_urban ** | 0.061 | 0.168 | −0.284 | <0.001 | 0.738 | <0.001 | −0.081 | 0.698 |
| Tele_speed_mean_urban | −0.016 | <0.001 | −0.016 | <0.001 | −0.015 | 0.054 | −0.017 | 0.170 |
| Tele_speed_max_highway | 0.004 | <0.001 | >−0.001 | 0.987 | 0.007 | 0.012 | −0.007 | 0.136 |
| $\sigma_\tau$ | 0.183 | <0.001 | - | - | 1.141 | 0.123 | - | - |
| AIC | 70,717.74 | | 34,891.44 | | 6901.30 | | 1890.13 | |

* we consider either all claims or only at fault TPL claims, correspondingly; ** expressed in thousands of kilometres.

**Table A10.** Cross-table with the parameter estimates of random effects and fixed-effects Poisson panel data models, with *p*-values. for the total number of claims (left) and at-fault third party liability (TPL) claims (right) per month and per insured based on the seven observed months in 2018 and 2019. Interactions are included.

| Variable | All Claims | | | | Only At Fault TPL Claims | | | |
|---|---|---|---|---|---|---|---|---|
| | Random Effects | | Fixed Effects | | Random Effects | | Fixed Effects | |
| | Parameter | *p*-Value | Parameter | *p*-Value | Parameter | *p*-Value | Parameter | *p*-Value |
| Intercept | −3.342 | <0.001 | - | - | −6.197 | 0.009 | - | - |
| lag(claims *) | −0.445 | <0.001 | −0.569 | <0.001 | −1.203 | 0.251 | −4.370 | <0.001 |
| Wind | 0.061 | 0.295 | 0.015 | 0.809 | −0.011 | 0.961 | −0.042 | 0.872 |
| Tele_km_total_urban ** | 0.083 | 0.752 | −0.335 | 0.244 | 0.847 | 0.253 | −0.983 | 0.375 |
| Tele_speed_mean_urban | −0.007 | 0.609 | <0.001 | 0.977 | −0.067 | 0.183 | −0.061 | 0.295 |
| Tele_speed_max_highway | 0.005 | 0.295 | −0.004 | 0.419 | 0.005 | 0.762 | −0.012 | 0.551 |
| Wind*Tele_km_total_urban ** | 0.004 | 0.883 | 0.011 | 0.678 | −0.005 | 0.941 | 0.103 | 0.324 |
| Wind*Tele_speed_mean_urban | −0.001 | 0.291 | −0.002 | 0.143 | 0.004 | 0.388 | 0.003 | 0.563 |
| Wind*Tele_speed_max_highway | >−0.001 | 0.799 | <0.001 | 0.538 | <0.001 | 0.922 | <0.001 | 0.931 |
| $\sigma_\tau$ | 0.158 | <0.001 | - | - | 1.006 | 0.161 | - | - |
| AIC | 58,916.38 | | 26,837.19 | | 5784.357 | | 1426.86 | |

* we consider either all claims or only at fault TPL claims, correspondingly; ** expressed in thousands of kilometres.

**Table A11.** Cross-table with the parameter estimates of random effects and fixed-effects Poisson panel data models, with *p*-values, for the total number of claims (left) and at-fault third party liability (TPL) claims (right) per month and per insured based on the seven observed months in 2018 and 2019. The models consider the average number of sunshine hours per day as explanatory variable describing weather conditions. Interactions are included.

| Variable | All Claims | | | | Only at Fault TPL Claims | | | |
|---|---|---|---|---|---|---|---|---|
| | Random Effects | | Fixed Effects | | Random Effects | | Fixed Effects | |
| | Parameter | *p*-Value | Parameter | *p*-Value | Parameter | *p*-Value | Parameter | *p*-Value |
| Intercept | −3.655 | <0.001 | - | - | −7.481 | <0.001 | - | - |
| lag(claims *) | −0.445 | <0.001 | −0.569 | <0.001 | −1.207 | 0.250 | −4.366 | <0.001 |
| Sun | 0.120 | 0.004 | 0.141 | 0.001 | 0.156 | 0.350 | 0.377 | 0.047 |
| Tele_km_total_urban ** | −0.366 | 0.012 | −0.717 | <0.001 | 0.336 | 0.468 | −0.759 | 0.239 |
| Tele_speed_mean_urban | −0.002 | 0.807 | 0.002 | 0.833 | 0.005 | 0.872 | 0.030 | 0.397 |
| Tele_speed_max_highway | 0.009 | <0.001 | 0.005 | 0.059 | 0.014 | 0.157 | 0.005 | 0.644 |
| Sun*Tele_km_total_urban ** | 0.059 | <0.001 | 0.061 | 0.001 | 0.055 | 0.306 | 0.093 | 0.205 |
| Sun*Tele_speed_mean_urban | −0.002 | 0.012 | −0.003 | 0.005 | −0.003 | 0.350 | −0.007 | 0.083 |
| Sun*Tele_speed_max_highway | −0.001 | 0.013 | −0.001 | 0.009 | −0.001 | 0.389 | −0.002 | 0.123 |
| $\sigma_\tau$ | 0.158 | <0.001 | - | - | 0.970 | 0.151 | - | - |
| AIC | 58,904.37 | | 26,818.82 | | 5791.49 | | 1426.45 | |

* we consider either all claims or only at fault TPL claims, correspondingly; ** expressed in thousands of kilometres.

**Table A12.** Cross-table with the parameter estimates of random effects and fixed-effects Poisson panel data models, with *p*-values, for the total number of claims (left) and at-fault third party liability (TPL) claims (right) per month and per insured based on the seven observed months in 2018 and 2019. The models consider the average monthly temperature as explanatory variable describing weather conditions. Interactions are included.

| Variable | All Claims | | | | Only at Fault TPL Claims | | | |
|---|---|---|---|---|---|---|---|---|
| | Random Effects | | Fixed Effects | | Random Effects | | Fixed Effects | |
| | Parameter | *p*-Value | Parameter | *p*-Value | Parameter | *p*-Value | Parameter | *p*-Value |
| Intercept | −2.952 | <0.001 | - | - | −6.480 | <0.001 | - | - |
| lag(claims *) | −0.445 | <0.001 | −0.570 | <0.001 | −1.190 | 0.256 | −4.367 | <0.001 |
| Temperature | 0.015 | 0.358 | 0.022 | 0.201 | 0.016 | 0.812 | 0.090 | 0.236 |
| Tele_km_total_urban ** | −0.071 | 0.573 | −0.460 | 0.001 | 0.650 | 0.102 | −0.494 | 0.362 |
| Tele_speed_mean_urban | −0.003 | 0.640 | −0.002 | 0.764 | 0.022 | 0.372 | 0.039 | 0.194 |
| Tele_speed_max_highway | 0.003 | 0.123 | −0.001 | 0.758 | 0.003 | 0.720 | −0.008 | 0.400 |
| Temperature*Tele_km_total_urban ** | 0.011 | 0.111 | 0.014 | 0.054 | 0.008 | 0.711 | 0.029 | 0.301 |
| Temperature*Tele_speed_mean_urban | −0.001 | 0.005 | −0.001 | 0.005 | −0.003 | 0.061 | −0.004 | 0.014 |
| Temperature*Tele_speed_max_highway | <0.001 | 0.992 | >−0.001 | 0.795 | <0.001 | 0.740 | >−0.001 | 0.743 |
| $\sigma_\tau$ | 0.158 | <0.001 | - | - | 0.988 | 0.157 | - | - |
| AIC | 58,916.62 | | 26,831.48 | | 5789.44 | | 1425.98 | |

* we consider either all claims or only at fault TPL claims, correspondingly; ** expressed in thousands of kilometres.

## References

Abdel-Aty, Mohamed A., and A. Essam Radwan. 2000. Modeling traffic accident occurrence and involvement. *Accident Analysis & Prevention* 32: 633–42. [CrossRef]

Agencia Estatal de Meteorología. 2022. Datos y Estadísticas. Servicio del Banco de Datos Nacional de Climatología. Available online: https://www.aemet.es/es/lineas_de_interes/datos_y_estadistica (accessed on 10 January 2023).

Ayuso, Mercedes, Montserrat Guillen, and Ana Maria Pérez-Marín. 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. [CrossRef] [PubMed]

Barry, Laurence, and Arthur Charpentier. 2020. Personalization as a promise: Can Big Data change the practice of insurance? *Big Data & Society* 7: 2053951720935143. [CrossRef]

Blier-Wong, Christopher, Helena Cossette, Luc Lamontagne, and Etienne Marceau. 2020. Machine learning in P&C insurance: A review for pricing and reserving. *Risks* 9: 4. [CrossRef]

Boucher, Jean-Philippe, Ana Maria Pérez-Marín, and Miguel Santolino. 2013. Pay-as-you-drive insurance: The effect of the kilometers on the risk of accident. *Anales Del Instituto de Actuarios Españoles* 19: 135–54. Available online: https://www.researchgate.net/profile/Miguel-Santolino/publication/285087799_Pay-As-You-Drive_Insurance_The_Effect_of_The_Kilometers_on_the_Risk_of_Accident/links/59bf7c3c0f7e9b48a29b5c80/Pay-As-You-Drive-Insurance-The-Effect-of-The-Kilometers-on-the-Risk-of-Accident.pdf (accessed on 5 March 2023).

Boucher, Jean-Philippe, and Roxane Turcotte. 2020. A longitudinal analysis of the impact of distance driven on the probability of car accidents. *Risks* 8: 91. [CrossRef]

Boucher, Jean-Philippe, Steven Côté, and Montserrat Guillen. 2017. Exposure as duration and distance in telematics motor insurance using generalized additive models. *Risks* 5: 54. [CrossRef]

Cabrera-Arnau, Carmen, Rafael Prieto Curiel, and Steven Richard Bishop. 2020. Uncovering the behaviour of road accidents in urban areas. *Royal Society Open Science* 7: 191739. [CrossRef]

Chan, Jennifer S., S. Boris Choy, Udi Makov, Ariel Shamir, and Vered Shapovalov. 2022. Variable Selection Algorithm for a Mixture of Poisson Regression for Handling Overdispersion in Claims Frequency Modeling Using Telematics Car Driving Data. *Risks* 10: 83. [CrossRef]

Che, Xin, Andre Liebenberg, and Jianren Xu. 2021. Usage-Based Insurance—Impact on Insurers and Potential Implications for InsurTech. *North American Actuarial Journal* 26: 428–55. [CrossRef]

Cheng, Jiang, Frank Y. Feng, and Xudong Zeng. 2022. Pay-As-You-Drive Insurance: Modeling and Implications. *North American Actuarial Journal* 1–19. [CrossRef]

Corradin, Alexandre, Michel Denuit, Marcin Detyniecki, Vincent Grari, Matteo Sammarco, and Julien Trufin. 2022. Joint modeling of claim frequencies and behavioral signals in motor insurance. *ASTIN Bulletin: The Journal of the IAA* 52: 33–54. [CrossRef]

Croissant, Yves. 2021. pglm: Panel Generalized Linear Models. Available online: https://cran.r-project.org/web/packages/pglm/index.html (accessed on 28 February 2023).

Duval, Francis, Jean Philippe Boucher, and Mathieu Pigeon. 2022. How much telematics information do insurers need for claim classification? *North American Actuarial Journal* 26: 570–90. [CrossRef]

Eling, Martin, and Mirko Kraft. 2020. The impact of telematics on the insurability of risks. *Journal of Risk Finance* 21: 77–109. [CrossRef]

Frees, Edward W., and Fei Huang. 2021. The Discriminating (Pricing) Actuary. *North American Actuarial Journal*. [CrossRef]

Gao, Guangyuan, Shengwang Meng, and Mario V. Wüthrich. 2022. What can we learn from telematics car driving data: A survey. *Insurance: Mathematics and Economics* 104: 185–99. [CrossRef]

Gao, Lisa, and Peng Shi. 2022. Leveraging high-resolution weather information to predict hail damage claims: A spatial point process for replicated point patterns. *Insurance: Mathematics and Economics* 107: 161–79. [CrossRef]

Guillen, Montserrat, Jens Perch Nielsen, Ana Maria Pérez-Marín, and Valandis Elpidorou. 2020. Can automobile insurance telematics predict the risk of near-miss events? *North American Actuarial Journal* 24: 141–52. [CrossRef]

Guillen, Montserrat, Jens Perch Nielsen, Mercedes Ayuso, and Ana Maria Pérez-Marín. 2019. The use of telematics devices to improve automobile insurance rates. *Risk Analysis* 39: 662–72. [CrossRef] [PubMed]

Huang, Yifan, and Shengwang Meng. 2019. Automobile insurance classification ratemaking based on telematics driving data. *Decision Support Systems* 127: 113156. [CrossRef]

Liang, Mingming, Dongdong Zhao, Yile Wu, Pengpeng Ye, Yuan Wang, Zhenhai Yao, Peng Bi, Leilei Duan, and Ye Sun. 2021. Short-term effects of ambient temperature and road traffic accident injuries in Dalian, Northern China: A distributed lag non-linear analysis. *Accident Analysis & Prevention* 153: 106057. [CrossRef]

Ma, Yu Luen, Xiaoyu Zhu, Xianbiao Hu, and Yi Chang Chiu. 2018. The use of context-sensitive insurance telematics data in auto insurance rate making. *Transportation Research Part A: Policy and Practice* 113: 243–58. [CrossRef]

Malin, Fanny, Ilkka Norros, and Satu Innamaa. 2019. Accident risk of road and weather conditions on different road types. *Accident Analysis and Prevention* 122: 181–88. [CrossRef] [PubMed]

Meng, Shengwang, He Wang, Yanlin Shi, and Guangyuan Gao. 2022. Improving automobile insurance claims frequency prediction with telematics car driving data. *ASTIN Bulletin: The Journal of the IAA* 52: 363–91. [CrossRef]

Mornet, Alexandre, Thomas Opitz, Michel Luzi, and Stephane Loisel. 2015. Index for Predicting Insurance Claims from Wind Storms with an Application in France. *Risk Analysis* 35: 2029–56. [CrossRef]

Owens, Emer, Barry Sheehan, Martin Mullins, Martin Cunneen, Juliane Ressel, and German Castignani. 2022. Explainable Artificial Intelligence (XAI) in Insurance. *Risks* 10: 230. [CrossRef]

Pérez-Marín, Ana Maria, and Montserrat Guillen. 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. [CrossRef]

Pérez-Marín, Ana Maria, Montserrat Guillen, Manuela Alcañiz, and Lluis Bermúdez. 2019. Quantile regression with telematics information to assess the risk of driving above the posted speed limit. *Risks* 7: 80. [CrossRef]

Pitarque, Albert, and Montserrat Guillen. 2022. Interpolation of quantile regression to estimate driver's risk of traffic accident based on excess speed. *Risks* 10: 19. [CrossRef]

Qiu, Lin, and Wilfred A. Nixon. 2008. Effects of Adverse Weather on Traffic Crashes: Systematic Review and Meta-Analysis. *Transportation Research Record* 2055: 139–46. [CrossRef]

So, Banghee, Jean Philippe Boucher, and Emiliano A. Valdez. 2021. Cost-sensitive multi-class adaboost for understanding driving behavior based on telematics. *ASTIN Bulletin: The Journal of the IAA* 51: 719–51. [CrossRef]

Williams, Allen R., Yoolim Jin, Anthony Duer, Tuka Alhani, and Mohammed Ghassemi. 2022. Nightly Automobile Claims Prediction from Telematics-Derived Features: A Multilevel Approach. *Risks* 10: 118. [CrossRef]

Zwerling, Craig, Corinne Peek-Asa, Paul S. Whitten, Seong Woo Choi, Nancy L. Sprince, and Michael P. Jones. 2005. Fatal motor vehicle crashes in rural and urban areas: Decomposing rates into contributing factors. *Injury Prevention* 11: 24–28. [CrossRef] [PubMed]