Driving data for automobile insurance: will telematics change ratemaking?

Montserrat Guillén

University of Barcelona
mguillen@ub.edu
www ub edu / riskcenter

SAV, Lucerne, Friday 30 August 2019
1. Introduction

2. Transition to telematics motor insurance

3. Data and results

4. Going forward to optimal pricing
1 Introduction
How do telematics data look like?

Sample Trip Summary Data – One Day

<table>
<thead>
<tr>
<th>State Date</th>
<th>Start Time</th>
<th>Motorway Yards</th>
<th>Urban Yards</th>
<th>Other Yards</th>
<th>Motorway Seconds</th>
<th>Urban Seconds</th>
<th>Other Seconds</th>
<th>Total Speeding Yards</th>
<th>Total Speeding Seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td>3/3/2012</td>
<td>12:12:00</td>
<td>-</td>
<td>31</td>
<td>-</td>
<td>-</td>
<td>13,713</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3/3/2012</td>
<td>14:17:11</td>
<td>-</td>
<td>3,355</td>
<td>-</td>
<td>-</td>
<td>7,934</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3/3/2012</td>
<td>14:34:03</td>
<td>39,566</td>
<td>39,010</td>
<td>69,042</td>
<td>1,328</td>
<td>1,922</td>
<td>2,864</td>
<td>1,379</td>
<td>38</td>
</tr>
<tr>
<td>3/3/2012</td>
<td>15:47:59</td>
<td>-</td>
<td>11,346</td>
<td>907</td>
<td>-</td>
<td>858</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3/3/2012</td>
<td>17:21:11</td>
<td>31,426</td>
<td>43,634</td>
<td>57,937</td>
<td>1,020</td>
<td>2,141</td>
<td>6,916</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3/3/2012</td>
<td>19:36:07</td>
<td>-</td>
<td>4,501</td>
<td>5,401</td>
<td>-</td>
<td>2,912</td>
<td>330</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3/3/2012</td>
<td>21:57:27</td>
<td>-</td>
<td>14,255</td>
<td>1,394</td>
<td>-</td>
<td>22,466</td>
<td>60</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3/3/2012</td>
<td>22:24:43</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>386</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Source: Jim Janavich ideas.returnonintelligence.com

2.6 MB per vehicle per week × 52 weeks × 50,000 vehicles = 6.8 TB per year
Companies selling motor insurance based on telematics

Pay-by-mile car insurance for savvy drivers.

A simple, straightforward policy that better fits the way you live.
Main questions

- Should **pay-per-mile** replace traditional motor insurance? **No**

- Will **telematics** transform motor insurance pricing? **Yes**

- What detailed **telematics data** should be collected? **Only valuable**
Main questions

- Should **pay-per-mile** replace traditional motor insurance? **No**

- Will **telematics** transform motor insurance pricing? **Yes**

- What detailed **telematics data** should be collected? **Only valuable**
Main questions

- Should pay-per-mile replace traditional motor insurance? No

- Will telematics transform motor insurance pricing? Yes

- What detailed telematics data should be collected? Only valuable
Introduction

Transition to telematics

Data and results

Going forward to optimal pricing

Basic concepts

- **Usage-Based-Insurance (UBI)**. Telemetry provides the insurer with detailed information on the use of the vehicle and the **premium is calculated based on usage**.
  - Pay-As-You-Drive (PAYD) automobile insurance is a policy agreement linked to vehicle driven distance.
  - Pay-How-You-Drive (PHYD) considers driving patterns.
• **Usage-Based-Insurance (UBI).** Telemetry provides the insurer with detailed information on the use of the vehicle and the \textbf{premium is calculated based on usage.}

  • **Pay-As-You-Drive (PAYD)** automobile insurance is a policy agreement linked to \textbf{vehicle driven distance.}

  • **Pay-How-You-Drive (PHYD)** considers \textbf{driving patterns.}
• **Usage-Based-Insurance (UBI).** Telemetry provides the insurer with detailed information on the use of the vehicle and the **premium is calculated based on usage.**
  • **Pay-As-You-Drive (PAYD)** automobile insurance is a policy agreement linked to **vehicle driven distance.**
  • **Pay-How-You-Drive (PHYD)** considers **driving patterns.**
The relationship between the distance run by a vehicle and the risk of accident has been discussed by many authors, most of them arguing that this relationship is not proportional (Litman, 2005 and 2011; Langford et al., 2008; Boucher et al., 2013).

There is evidence of the relationship between speed, type of road, urban and night-time driving and the risk of accident (Rice et al., 2003; Laurie, 2011; Ellison et al, 2015; Wüthrich, 2017; Verbelen et al. 2018; Ma et al. 2018; Gao, Yang and Wüthrich, 2019).

Telematics information can replace some traditional rating factors and provide a pricing model with the same predictive performance (Verbelen et al. 2018; Ayuso et al., 2016b; Baecke and Bocca, 2017).

Gender: discrimination that turns out to be a proxy

Gender can be replaced by:

- km/day (Barcelona approach)
- km/trip (Leuven approach)
The relationship between the distance run by a vehicle and the risk of accident has been discussed by many authors, most of them arguing that this relationship is not proportional (Litman, 2005 and 2011; Langford et al., 2008; Boucher et al., 2013).

There is evidence of the relationship between speed, type of road, urban and night-time driving and the risk of accident (Rice et al., 2003; Laurie, 2011; Ellison et al, 2015; Wüthrich, 2017; Verbelen et al. 2018; Ma et al. 2018; Gao, Yang and Wüthrich, 2019).

Telematics information can replace some traditional rating factors and provide a pricing model with the same predictive performance (Verbelen et al. 2018; Ayuso et al., 2016b; Baecke and Bocca, 2017).

Gender: discrimination that turns out to be a proxy

Gender can be replaced by:

- km/day (Barcelona approach)
- km/trip (Leuven approach)
The relationship between the distance run by a vehicle and the risk of accident has been discussed by many authors, most of them arguing that this relationship is not proportional (Litman, 2005 and 2011; Langford et al., 2008; Boucher et al., 2013).

There is evidence of the relationship between speed, type of road, urban and night-time driving and the risk of accident (Rice et al., 2003; Laurie, 2011; Ellison et al, 2015; Wüthrich, 2017; Verbelen et al. 2018; Ma et al. 2018; Gao, Yang and Wüthrich, 2019).

Telematics information can replace some traditional rating factors and provide a pricing model with the same predictive performance (Verbelen et al. 2018; Ayuso et al., 2016b; Baecke and Bocca, 2017).

Gender: discrimination that turns out to be a proxy

Gender can be replaced by:

km/day (Barcelona approach) or km/trip (Leuven approach)
1. The relationship between the distance run by a vehicle and the risk of accident has been discussed by many authors, most of them arguing that this relationship is not proportional (Litman, 2005 and 2011; Langford et al., 2008; Boucher et al., 2013).

2. There is evidence of the relationship between speed, type of road, urban and night-time driving and the risk of accident (Rice et al., 2003; Laurie, 2011; Ellison et al, 2015; Wüthrich, 2017; Verbelen et al. 2018; Ma et al. 2018; Gao, Yang and Wüthrich, 2019).

3. Telematics information can replace some traditional rating factors and provide a pricing model with the same predictive performance (Verbelen et al. 2018; Ayuso et al., 2016b; Baecke and Bocca, 2017).

Gender: discrimination that turns out to be a proxy

Gender can be replaced by: km/day (Barcelona approach) or km/trip (Leuven approach)
Strong evidence exists

- Information on mileage and driving habits improves the prediction of the number of claims (and the cost of claims) compared to traditional rating factors and coverage exclusively by time (usually one year).

- Semi-autonomous vehicles are expected to contribute to a lower frequency of motor accidents

Our question is:

What is the role of telematics in motor insurance ratemaking?
Strong evidence exists

- Information on mileage and driving habits improves the prediction of the number of claims (and the cost of claims) compared to traditional rating factors and coverage exclusively by time (usually one year).
- Semi-autonomous vehicles are expected to contribute to a lower frequency of motor accidents

Our question is:

What is the role of telematics in motor insurance ratemaking?
Strong evidence exists

- Information on mileage and driving habits improves the prediction of the number of claims (and the cost of claims) compared to traditional rating factors and coverage exclusively by time (usually one year).

- Semi-autonomous vehicles are expected to contribute to a lower frequency of motor accidents

Our question is:

What is the role of telematics in motor insurance ratemaking?
**Strong evidence exists**

- Information on mileage and driving habits improves the prediction of the **number of claims** (and the **cost of claims**) compared to traditional rating factors and coverage exclusively by time (usually one year).

- Semi-autonomous vehicles are expected to contribute to a **lower frequency of motor accidents**

**Our question is:**

What is the role of telematics in motor insurance ratemaking?
Impact:

- **Distance driven** (mileage, exposure to risk) and other telematics data (speed, braking, habits) modify traditional premium calculation.

Our contribution:

- Propose a method to **update premiums** regularly with telematics data. We create the basis for real-time pricing (not necessary), and real-time prevention.

- Show that the price per mile depends on driving habits and price should not be proportional to distance driven. A zero claim is relatively more frequent for intensive users. Propose a predictive modeling approach for this purpose.

- Derive some **open-questions** about risk measures to summarize telematics big data and optimal pricing when customers may lapse.
Impact:

- **Distance driven** (mileage, exposure to risk) and **other telematics data** (speed, braking, habits) modify traditional premium calculation.

Our contribution:

- Propose a method to **update premiums** regularly with telematics data. We create the basis for **real-time pricing** (not necessary), and **real-time prevention**.

- Show that the **price per mile depends on driving habits** and price should not be proportional to distance driven. A zero claim is relatively more frequent for intensive users. Propose a predictive modeling approach for this purpose.

- Derive some **open-questions** about risk measures to summarize telematics big data and **optimal pricing** when customers may lapse.
Impact:

- **Distance driven** (mileage, exposure to risk) and **other telematics data** (speed, braking, habits) modify traditional premium calculation.

Our contribution:

- Propose a method to **update premiums** regularly with telematics data. We create the basis for **real-time pricing** (not necessary), and **real-time prevention**.

- Show that the **price per mile depends on driving habits** and price should not be proportional to distance driven. A zero claim is relatively more frequent for intensive users. Propose a predictive modeling approach for this purpose.

- Derive some **open-questions** about risk measures to summarize telematics big data and **optimal pricing** when customers may lapse.
Impact:

- **Distance driven** (mileage, exposure to risk) and **other telematics data** (speed, braking, habits) modify traditional premium calculation.

Our contribution:

- Propose a method to **update premiums** regularly with telematics data. We create the basis for **real-time pricing** (not necessary), and **real-time prevention**.
- Show that the **price per mile depends on driving habits** and price should not be proportional to distance driven. A zero claim is relatively more frequent for intensive users. Propose a predictive modeling approach for this purpose.
- Derive some **open-questions** about risk measures to summarize telematics big data and **optimal pricing** when customers may lapse.
Impact:

- Distance driven (mileage, exposure to risk) and other telematics data (speed, braking, habits) modify traditional premium calculation.

Our contribution:

- Propose a method to update premiums regularly with telematics data. We create the basis for real-time pricing (not necessary), and real-time prevention.

- Show that the price per mile depends on driving habits and price should not be proportional to distance driven. A zero claim is relatively more frequent for intensive users. Propose a predictive modeling approach for this purpose.

- Derive some open-questions about risk measures to summarize telematics big data and optimal pricing when customers may lapse.
Introduction

Transition to telematics

Data and results

Going forward to optimal pricing

What we do

Impact:

- **Distance driven** (mileage, exposure to risk) and **other telematics data** (speed, braking, habits) modify traditional premium calculation.

Our contribution:

- Propose a method to **update premiums** regularly with telematics data. We create the basis for **real-time pricing** (not necessary), and **real-time prevention**.

- Show that the **price per mile depends on driving habits** and price should not be proportional to distance driven. A zero claim is relatively more frequent for intensive users. Propose a predictive modeling approach for this purpose.

- Derive some **open-questions** about risk measures to summarize **telematics big data** and **optimal pricing** when customers may lapse.
Why is insurance analytics a good example of big data in applied economics?

Source: Guillen, 2016
2 Transition to telematics
The classical ratemaking model is based on a prediction of the number of claims (usually for one year) times the average claim cost plus some extra loadings.

- Subscript $i$ denotes the $i$th policy holder in a portfolio of $n$ insureds.
- Given $x_i = (x_{i1}, \ldots, x_{ik})$ (vector of $k$ covariates), the number of claims $Y_i$ (dependent variable) follows a Poisson distribution with parameter $\lambda_i$, which is a function of the linear combination of parameters and regressors, $\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}$.

$$E(Y_i|x_i) = \exp(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}) \quad (1)$$

The unknown parameters to be estimated are $(\beta_0, \ldots, \beta_k)$.
- Classical covariates are age, time since driver’s license was issued, driving zone, type of car, ...
- The pure premium equals the product of the expected number of claims times the average claim cost. Finally, the premium is obtained once additional margins and safety loadings are included.
The **classical ratemaking** model is based on a prediction of the number of claims (usually for one year) times the average claim cost plus some extra loadings.

- **Subscript** \( i \) denotes the \( i \)th policy holder in a portfolio of \( n \) insureds.

- Given \( x_i = (x_{1i}, \ldots, x_{ki}) \) (vector of \( k \) covariates), the number of claims \( Y_i \) (dependent variable) follows a Poisson distribution with parameter \( \lambda_i \), which is a function of the linear combination of parameters and regressors, \( \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} \).

\[
E(Y_i|x_i) = \exp(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}) \tag{1}
\]

The unkown parameters to be estimated are \( (\beta_0, \ldots, \beta_k) \).

- Classical covariates are age, time since driver’s license was issued, driving zone, type of car,...

- The **pure premium** equals the product of the expected number of claims times the average claim cost. Finally, the **premium** is obtained once additional margins and safety loadings are included.
The **classical ratemaking** model is based on a prediction of the number of claims (usually for one year) times the average claim cost plus some extra loadings.

- Subscript $i$ denotes the $i$th policy holder in a portfolio of $n$ insureds.
- Given $x_i = (x_{1i}, \ldots, x_{ki})$ (vector of $k$ covariates), the number of claims $Y_i$ (dependent variable) follows a Poisson distribution with parameter $\lambda_i$, which is a function of the linear combination of parameters and regressors, $\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}$.

$$E(Y_i|x_i) = \exp(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}) \quad (1)$$

The unknown parameters to be estimated are $(\beta_0, \ldots, \beta_k)$.

- Classical covariates are age, time since driver’s license was issued, driving zone, type of car,...
- The **pure premium** equals the product of the expected number of claims times the average claim cost. Finally, the **premium** is obtained once additional margins and safety loadings are included.
The classical ratemaking model is based on a prediction of the number of claims (usually for one year) times the average claim cost plus some extra loadings.

- Subscript $i$ denotes the $i$th policy holder in a portfolio of $n$ insureds.
- Given $x_i = (x_{1i}, \ldots x_{ki})$ (vector of $k$ covariates), the number of claims $Y_i$ (dependent variable) follows a Poisson distribution with parameter $\lambda_i$, which is a function of the linear combination of parameters and regressors, $\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}$.

$$E(Y_i|x_i) = \exp(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik})$$ (1)

The unknown parameters to be estimated are $(\beta_0, \ldots, \beta_k)$.

- Classical covariates are age, time since driver’s license was issued, driving zone, type of car, ...
- The pure premium equals the product of the expected number of claims times the average claim cost. Finally, the premium is obtained once additional margins and safety loadings are included.
The classical ratemaking model is based on a prediction of the number of claims (usually for one year) times the average claim cost plus some extra loadings.

- Subscript $i$ denotes the $i$th policy holder in a portfolio of $n$ insureds.
- Given $x_i = (x_{1i}, ... x_{ki})$ (vector of $k$ covariates), the number of claims $Y_i$ (dependent variable) follows a Poisson distribution with parameter $\lambda_i$, which is a function of the linear combination of parameters and regressors, $\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}$.

$$E(Y_i|x_i) = \exp(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}) \quad (1)$$

The unknown parameters to be estimated are $(\beta_0, \ldots, \beta_k)$.

- Classical covariates are age, time since driver’s license was issued, driving zone, type of car, ...
- The pure premium equals the product of the expected number of claims times the average claim cost. Finally, the premium is obtained once additional margins and safety loadings are included.
Improving automobile insurance ratemaking using telematics: incorporating mileage and driver behaviour data

Mercedes Ayuso\textsuperscript{1} \& Montserrat Guillen\textsuperscript{1} \& Jens Perch Nielsen\textsuperscript{2}

\textsuperscript{1}Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract We show how data collected from a GPS device can be incorporated in motor insurance ratemaking. The calculation of premium rates based upon driver behaviour represents an opportunity for the insurance sector. Our approach is based on count data regression models for frequency, where exposure is driven by the distance travelled and additional parameters that capture characteristics of automobile usage and which may affect claiming behaviour. We propose implementing a classical frequency model that is updated with telemetrics information. We illustrate the method using real data from usage-based insurance policies. Results show that not only the distance travelled by the driver, but also driver habits, significantly influence the expected number of accidents and, hence, the cost of insurance coverage. This paper provides a methodology including a transition pricing transferring knowledge and experience that the company already had before the telematics data arrived to the new world including telematics information.

Keywords Tariff \& Premium calculation \& Pay-as-you-drive insurance \& Count data models
In *Transportation* (2018) we proposed a method for assessing the influence on the expected frequency of usage-based variables which can be viewed as a **correction of the classical ratemaking model**.

A two-step procedure:

- Step 1: Let \( \hat{Y}_i \) be the frequency estimate obtained as a function of the classical explanatory covariates \( x_i = (x_{i1}, \ldots, x_{ik}) \).
- Step 2: Let \( z_i = (z_{i1}, \ldots, z_{il}) \) be the information collected periodically from a telematics unit. Then, the prediction from usage-based insurance information is a correction such that:

\[
E(Y_{iUBI}|z_i, \hat{Y}_i) = \hat{Y}_i \exp(\eta_0 + \eta_1 z_{i1} + \ldots + \eta_k z_{ik}),
\]

where the parameter estimates \((\eta_0, \ldots, \eta_l)\) can now be obtained using \( \hat{Y}_i \) as an offset.

**Note:**

This approach is less efficient than a full information model, but it works well in practice. Telematics data are collected on a continuous basis and this correction can be implemented regularly (i.e. on a weekly basis).
In *Transportation* (2018) we proposed a method for assessing the influence on the expected frequency of usage-based variables which can be viewed as a **correction of the classical ratemaking model**.

A two-step procedure:

- **Step 1**: Let $\hat{Y}_i$ be the frequency estimate obtained as a function of the classical explanatory covariates $x_i = (x_{i1}, \ldots, x_{ik})$.

- **Step 2**: Let $z_i = (z_{i1}, \ldots, z_{il})$ be the information collected periodically from a telematics unit. Then, the prediction from usage-based insurance information is a correction such that:

$$E(Y_{iUBI}^{|z_i, \hat{Y}_i}) = \hat{Y}_i \exp(\eta_0 + \eta_1 z_{i1} + \ldots + \eta_k z_{ik}), \quad (2)$$

where the parameter estimates $(\eta_0, \ldots, \eta_l)$ can now be obtained using $\hat{Y}_i$ as an offset.

**Note:**

This approach is less efficient than a full information model, but it works well in practice. Telematics data are collected on a continuous basis and this correction can be implemented regularly (i.e. on a weekly basis).
In *Transportation* (2018) we proposed a method for assessing the influence on the expected frequency of usage-based variables which can be viewed as a **correction of the classical ratemaking model**.

A two-step procedure:

- **Step 1:** Let $\hat{Y}_i$ be the frequency estimate obtained as a function of the classical explanatory covariates $x_i = (x_{i1}, \ldots, x_{ik})$.

- **Step 2:** Let $z_i = (z_{i1}, \ldots, z_{il})$ be the information collected periodically from a telematics unit. Then, the prediction from usage-based insurance information is a correction such that:

$$E(Y_{iUBI}^{|z_i, \hat{Y}_i}) = \hat{Y}_i \exp(\eta_0 + \eta_1 z_{i1} + \ldots + \eta_k z_{ik}),$$

where the parameter estimates $(\eta_0, \ldots, \eta_l)$ can now be obtained using $\hat{Y}_i$ as an offset.

**Note:**

This approach is less efficient than a full information model, but it works well in practice. Telematics data are collected on a continuous basis and this correction can be implemented regularly (i.e. on a weekly basis).
In *Transportation* (2018) we proposed a method for assessing the influence on the expected frequency of usage-based variables which can be viewed as a **correction of the classical ratemaking model**. A two-step procedure:

- **Step 1:** Let $\hat{Y}_i$ be the frequency estimate obtained as a function of the classical explanatory covariates $x_i = (x_{i1}, \ldots, x_{ik})$.

- **Step 2:** Let $z_i = (z_{i1}, \ldots, z_{il})$ be the information collected periodically from a telematics unit. Then, the prediction from usage-based insurance information is a correction such that:

$$E(Y_{iUBI} | z_i, \hat{Y}_i) = \hat{Y}_i \exp(\eta_0 + \eta_1 z_{i1} + \ldots + \eta_k z_{ik}),$$

(2)

where the parameter estimates $(\eta_0, \ldots, \eta_l)$ can now be obtained using $\hat{Y}_i$ as an offset.

**Note:**

This approach is less efficient than a full information model, but it works well in practice. Telematics data are collected on a continuous basis and this correction can be implemented regularly (i.e. on a weekly basis).

Most automobile insurance databases contain a large number of policyholders with zero claims. This high frequency of zeros may reflect the fact that some insureds make little use of their vehicle, or that they do not wish to make a claim for small accidents in order to avoid an increase in their premium, but it might also be because of good driving. We analyse information on exposure to risk and driving habits using telematics data from a Pay-as-you-Drive sample of insureds. We include distance travelled per year as part of an offset in a zero-inflated Poisson model to predict the excess of zeros. We show the existence of a learning effect for large values of distance travelled, so that longer driving should result in higher premium, but there should be a discount for drivers that accumulate longer distances over time due to the increased proportion of zero claims. We confirm that speed limit violations and driving in urban areas increase the expected number of accident claims. We discuss how telematics information can be used to design better insurance and to improve traffic safety.
In **Risk Analysis** (2018) we propose to include the distance travelled per year as an offset in a Zero Inflated Poisson model to predict the number of claims in *Pay as You Drive* insurance.

- **The Poisson model with exposure:** Let us call $T_i$ the exposure factor for policy holder $i$, in our case $T_i = \ln(D_i)$, where $D_i$ indicates distance travelled, then:

$$E(Y_i|x_i, T_i) = D_i \exp(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}) = D_i \lambda_i \quad (3)$$

**Excess of zeros** exists because:

- Some insureds do not use their car and so they do not have claims
- Some insured acquire exceptionally good driving skills and they do not have claims (*learning curve*).
The Zero-inflated Poisson (ZIP) model: Now the probability of not suffering an accident is

\[ P(Y_i = 0) = p_i + (1 - p_i)P(Y^*_i = 0) \]  \hspace{1cm} (4)

where \( p_i \) is the probability of excess of zeros. \( Y^*_i \) follows a Poisson distribution with parameter \( \exp(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}) \), and \( p_i \) may depend on some covariates.
A ZIP Poisson model with exposure

We assume that \( p_i \) is the probability of an excess of zeros, and it is specified as a logistic regression model such that

\[
p_i = \frac{\exp(\alpha_0 + \alpha_1 \ln(D_i))}{1 + \exp(\alpha_0 + \alpha_1 \ln(D_i))}.
\]  

(5)

The Poisson model for \( Y^* \) is specified as follows, with an exposure

\[
E(Y_i^*|x_i, T_i) = D_i \exp(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}) = D_i \lambda_i = \exp(\ln(D_i)) \lambda_i = \exp(T_i) \lambda_i, \text{ where } T_i = \ln(D_i).
\]  

The expectation of the Poisson part is:

\[
(1 - p_i)E(Y_i^*|x_i, T_i) = \frac{1}{1 + \exp(\alpha_0 + \alpha_1 \ln(D_i))} D_i \lambda_i = D_i^* \lambda_i
\]  

(6)

where \( D_i^* = \frac{D_i}{1 + \exp(\alpha_0 + \alpha_1 \ln(D_i))} \) is a transformation of the original measure of exposure (distance driven) \( D_i \).
A **ZIP Poisson model with exposure**

So, when we include zero-inflation there is a transformation of the exposure in the Poisson part of the model.

- When $D_i$ is big then $D_i^* = \frac{D_i}{1 + \exp(\alpha_0 + \alpha_1 \ln(D_i))}$ tends to zero if $\alpha_1 > 1$.
- When $\alpha_1 = 1$ then $D_i^*$ tends to constant $\frac{1}{\exp(\alpha_0)}$ when $D_i$ increases.
- Assuming that $D_i \geq 1$, when $\alpha_1 > 1$ this is a concave transformation that scales exposure into the interval $\left[0, \frac{1}{1+\exp(\alpha_0)}\right]$. So, the larger the exposure the smaller the value whereas the smaller the exposure the larger the value.
- Assuming that $D_i \geq 1$, when $\alpha_1 \leq 1$ then the transformation is a change of scale to the interval $\left[\frac{1}{1+\exp(\alpha_0)}, +\infty\right)$.
A ZIP Poisson model with exposure

So, when we include zero-inflation there is a transformation of the exposure in the Poisson part of the model.

- When $D_i$ is big then
  $$D_i^* = \frac{D_i}{1 + \exp(\alpha_0 + \alpha_1 \ln(D_i))}$$
  tends to zero if $\alpha_1 > 1$.

- When $\alpha_1 = 1$ then $D_i^*$ tends to constant
  $$\frac{1}{\exp(\alpha_0)}$$
  when $D_i$ increases.

- Assuming that $D_i \geq 1$, when $\alpha_1 > 1$ this is a concave transformation that scales exposure into the interval $\left[0, \frac{1}{1 + \exp(\alpha_0)}\right]$. So, the larger the exposure the smaller the value whereas the smaller the exposure the larger the value.

- Assuming that $D_i \geq 1$, when $\alpha_1 \leq 1$ then the transformation is a change of scale to the interval $\left[\frac{1}{1 + \exp(\alpha_0)}, +\infty\right)$.
A *ZIP Poisson model with exposure*

So, when we include zero-inflation there is a transformation of the exposure in the Poisson part of the model.

- **When** $D_i$ **is big then** $D_i^* = \frac{D_i}{1+\exp(\alpha_0 + \alpha_1 \ln(D_i))}$ tend to zero if $\alpha_1 > 1$.
- **When** $\alpha_1 = 1$ then $D_i^*$ tends to constant $\frac{1}{\exp(\alpha_0)}$ when $D_i$ increases.
- **Assuming that** $D_i \geq 1$, **when** $\alpha_1 > 1$ this is a concave transformation that scales exposure into the interval $\left[0, \frac{1}{1+\exp(\alpha_0)}\right]$. So, the larger the exposure the smaller the value whereas the smaller the exposure the larger the value.
- **Assuming that** $D_i \geq 1$, **when** $\alpha_1 \leq 1$ then the transformation is a change of scale to the interval $\left[\frac{1}{1+\exp(\alpha_0)}, +\infty\right)$.
A ZIP Poisson model with exposure

So, when we include zero-inflation there is a transformation of the exposure in the Poisson part of the model.

- When $D_i$ is big then $D_i^* = \frac{D_i}{1+\exp(\alpha_0+\alpha_1 \ln(D_i))}$ tends to zero if $\alpha_1 > 1$.
- When $\alpha_1 = 1$ then $D_i^*$ tends to constant $\frac{1}{\exp(\alpha_0)}$ when $D_i$ increases.
- Assuming that $D_i \geq 1$, when $\alpha_1 > 1$ this is a concave transformation that scales exposure into the interval $\left[0, \frac{1}{1+\exp(\alpha_0)}\right]$. So, the larger the exposure the smaller the value whereas the smaller the exposure the larger the value.
- Assuming that $D_i \geq 1$, when $\alpha_1 \leq 1$ then the transformation is a change of scale to the interval $\left[\frac{1}{1+\exp(\alpha_0)}, +\infty\right)$. 
A ZIP Poisson model with exposure

If we look at the logistic regression part, we can also derive the following expression:

\[
p_i = \frac{\exp(\alpha_0 + \alpha_1 \ln(D_i))}{1 + \exp(\alpha_0 + \alpha_1 \ln(D_i))} = \frac{\exp(\alpha_0 + \alpha_1 \ln(D_i))}{1 + \exp(\alpha_0 + \alpha_1 \ln(D_i))} \frac{D_i}{D_i} = \exp(\alpha_0 + \alpha_1 \ln(D_i)) \frac{D_i^*}{D_i}
\]

So, the probability of zero excess \((p_i)\) can be understood as a rescaling of the relative transformed exposure.

Interestingly, when \(\alpha_1 < 0\) then note that \(p_i\) tends to zero when \(D_i\) increases, whereas when \(\alpha_1 > 0\) then \(p_i\) tends to one when \(D_i\) increases.

In the empirical part we find \(\alpha_1 > 0\), which means that there is a learning effect and the excess of zeros is more important than the Poisson part when distance driven increases.
Excessive braking or acceleration and other risky events

Can Automobile Insurance Telematics Predict the Risk of Near-Miss Events?

Montserrat Guillen, Jens Perch Nielsen, Ana M. Pérez-Marin, and Valandidis Elpidorou

Telematics data from usage-based motor insurance provide valuable information — including vehicle usage, attitude toward speeding, and time and proportion of urban/nonurban driving, which can be used for ratemaking. Additional information on acceleration, braking, and cornering can likewise be usefully employed to identify near-miss events, a concept taken from aviation that denotes a situation that might have resulted in an accident. We analyze near-miss events from a sample of drivers in order to identify the risk factors associated with a higher risk of near-miss occurrence. Our empirical application with a pilot sample of real usage-based insurance data reveals that certain factors are associated with a higher expected number of near-miss events, but that the association differs depending on the type of near miss. We conclude that nighttime driving is associated with a lower risk of cornering events, urban driving increases the risk of braking events, and speeding is associated with acceleration events. These results are relevant for the insurance industry in order to implement dynamic risk monitoring through telematics, as well as preventive actions.

1. INTRODUCTION AND MOTIVATION

Before the emergence of telematics, insurers had no verifiable information on the driving patterns and real vehicle usage of the insured. Driving circumstances and styles could only be determined, and then indirectly, in the specific case of an accident. Today, in contrast, telematics provides a novel source of data for risk classification before an accident, or even before a dangerous event, occurs, in what insurers refer to as “near-miss.” A near miss—a term taken from aviation safety, where reports
In *North American Actuarial Journal* (to appear 2020) we propose modelling near-miss events

- **Acceleration event** positive difference between the maximum acceleration reading and the acceleration detected in the first reading above the acceleration event detection threshold (set at 6m/s², see Hynes & Dickey, 2008).

- **Breaking event** same as acceleration, with a minus sign.

- **Cornering event** larger than one ratio between the speed of a reading and the maximum speed possible during a turn for the vehicle to stay on track.

We conclude that night-time driving is associated with a lower risk of cornering events, urban driving increases the risk of braking events and speeding is associated with acceleration events.

**Pricing versus safety**

Ethical question: should all drivers be penalized equally for each excessive near-miss event regardless of their driving zone?
In *North American Actuarial Journal* (to appear 2020) we propose modelling **near-miss events**

- **Acceleration event** positive difference between the maximum acceleration reading and the acceleration detected in the first reading above the acceleration event detection threshold (set at $6m/s^2$, see Hynes & Dickey, 2008).

- **Breaking event** same as acceleration, with a minus sign.

- **Cornering event** larger than one ratio between the speed of a reading and the maximum speed possible during a turn for the vehicle to stay on track.

We conclude that night-time driving is associated with a lower risk of cornering events, urban driving increases the risk of braking events and speeding is associated with acceleration events.

**Pricing versus safety**

Ethical question: should all drivers be penalized equally for each excessive near-miss event regardless of their driving zone?
In *North American Actuarial Journal* (to appear 2020) we propose modelling **near-miss events**

- **Acceleration event** positive difference between the maximum acceleration reading and the acceleration detected in the first reading above the acceleration event detection threshold (set at 6m/s², see Hynes & Dickey, 2008).
- **Breaking event** same as acceleration, with a minus sign.
- **Cornering event** larger than one ratio between the speed of a reading and the maximum speed possible during a turn for the vehicle to stay on track.

We conclude that night-time driving is associated with a lower risk of cornering events, urban driving increases the risk of braking events and speeding is associated with acceleration events.

**Pricing versus safety**

Ethical question: should all drivers be penalized equally for each excessive near-miss event regardless of their driving zone?
In North American Actuarial Journal (to appear 2020) we propose modelling near-miss events

- **Acceleration event** positive difference between the maximum acceleration reading and the acceleration detected in the first reading above the acceleration event detection threshold (set at $6\text{m/s}^2$, see Hynes & Dickey, 2008).

- **Breaking event** same as acceleration, with a minus sign.

- **Cornering event** larger than one ratio between the speed of a reading and the maximum speed possible during a turn for the vehicle to stay on track.

We conclude that night-time driving is associated with a lower risk of cornering events, urban driving increases the risk of braking events and speeding is associated with acceleration events.

Pricing versus safety

Ethical question: should all drivers be penalized equally for each excessive near-miss event regardless of their driving zone?
In *North American Actuarial Journal* (to appear 2020) we propose modelling **near-miss events**

- **Acceleration event** positive difference between the maximum acceleration reading and the acceleration detected in the first reading above the acceleration event detection threshold (set at 6m/s², see Hynes & Dickey, 2008).

- **Breaking event** same as acceleration, with a minus sign.

- **Cornering event** larger than one ratio between the speed of a reading and the maximum speed possible during a turn for the vehicle to stay on track.

We conclude that night-time driving is associated with a lower risk of **cornering events**, urban driving increases the risk of **braking events** and speeding is associated with **acceleration events**.

**Pricing versus safety**

Ethical question: should all drivers be penalized equally for each excessive near-miss event regardless of their driving zone?
3 Data and results
2009

MAPFRE, la aseguradora global de confianza
Information on the data sets
Introduction

Transition to telematics

Data and results

Going forward to optimal pricing

Information on the data sets
Zero-inflation for the Number of Claims
Empirical application based on 25,014 insureds with car insurance coverage throughout 2011, that is, individuals exposed to the risk for a full year.

Table I. Frequency of claims per driver (n=25,014) in the Spanish insurance dataset (all claims, at fault, and not at fault)

<table>
<thead>
<tr>
<th>Number of claims</th>
<th>All claims</th>
<th>Claims at fault</th>
<th>Claims not at fault</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>20,608</td>
<td>22,837</td>
<td>22,432</td>
</tr>
<tr>
<td>1</td>
<td>3,310</td>
<td>1,750</td>
<td>2,111</td>
</tr>
<tr>
<td>2</td>
<td>889</td>
<td>385</td>
<td>424</td>
</tr>
<tr>
<td>3</td>
<td>165</td>
<td>37</td>
<td>40</td>
</tr>
<tr>
<td>4</td>
<td>34</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

One insured driver had 6 claims, 2 were at fault and 4 where not at fault.

Table II. Descriptive statistics for the risk exposure indicator (total kilometres travelled per year in 000s)

<table>
<thead>
<tr>
<th></th>
<th>All Sample n = 25,014</th>
<th>Drivers with no claims n = 20,608 (82.4%)</th>
<th>Drivers with claims n = 4,406 (17.6%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>7.16</td>
<td>6.99</td>
<td>7.96</td>
</tr>
<tr>
<td>1st Quartile</td>
<td>4.14</td>
<td>4.00</td>
<td>4.87</td>
</tr>
<tr>
<td>Median</td>
<td>6.46</td>
<td>6.28</td>
<td>7.22</td>
</tr>
<tr>
<td>3rd Quartile</td>
<td>9.40</td>
<td>9.22</td>
<td>10.30</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>4.19</td>
<td>4.14</td>
<td>4.35</td>
</tr>
</tbody>
</table>
### Table 2  Descriptive statistics by claims (quantitative variables)

<table>
<thead>
<tr>
<th></th>
<th>All sample N = 25,014</th>
<th>Drivers with no claims N = 20,608 (82.4%)</th>
<th>Drivers with claims N = 4406 (17.6%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Age</td>
<td>27.57</td>
<td>3.09</td>
<td>27.65</td>
</tr>
<tr>
<td>Age driving licence</td>
<td>7.17</td>
<td>3.05</td>
<td>7.27</td>
</tr>
<tr>
<td>Vehicle age</td>
<td>8.75</td>
<td>4.17</td>
<td>8.76</td>
</tr>
<tr>
<td>Power</td>
<td>97.22</td>
<td>27.77</td>
<td>96.98</td>
</tr>
<tr>
<td>Km per year (000s)</td>
<td>7.16</td>
<td>4.19</td>
<td>6.99</td>
</tr>
<tr>
<td>Km per year at night (%)</td>
<td>6.91</td>
<td>6.35</td>
<td>6.85</td>
</tr>
<tr>
<td>Km per year over speed limit (%)</td>
<td>6.33</td>
<td>6.83</td>
<td>6.28</td>
</tr>
<tr>
<td>Urban km per year (%)</td>
<td>25.87</td>
<td>14.36</td>
<td>25.51</td>
</tr>
</tbody>
</table>

### Table 3  Descriptive statistics by claims (categorical variables)

<table>
<thead>
<tr>
<th></th>
<th>All sample N = 25,014</th>
<th>Drivers with no claims N = 20,608 (82.4%)</th>
<th>Drivers with claims N = 4406 (17.6%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Percent</td>
<td>Frequency</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>12,235</td>
<td>48.91</td>
<td>10,018</td>
</tr>
<tr>
<td>Women</td>
<td>12,779</td>
<td>51.09</td>
<td>10,590</td>
</tr>
<tr>
<td>Parking</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>19,356</td>
<td>77.38</td>
<td>15,912</td>
</tr>
<tr>
<td>No</td>
<td>5658</td>
<td>22.62</td>
<td>4696</td>
</tr>
</tbody>
</table>
Poisson model results. All types of claims.

**Table 6. Poisson model results with offset km per year. All claim types (n=25,014)**

<table>
<thead>
<tr>
<th></th>
<th>All variables</th>
<th>Non-telematics</th>
<th>Telematics</th>
<th>Telematics with offsets (Log of prediction of Non-telematics model - Column 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Coefficient</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intersect</td>
<td>-2.193</td>
<td>-0.472</td>
<td>-4.219</td>
<td>-0.731</td>
</tr>
<tr>
<td>Age</td>
<td>-0.145</td>
<td>-0.200</td>
<td>-0.009</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age^2</td>
<td>0.003</td>
<td>0.004</td>
<td>0.002</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Male</td>
<td>-0.086</td>
<td>-0.049</td>
<td>-0.731</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age Driving License</td>
<td>-0.061</td>
<td>-0.076</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Vehicle Age</td>
<td>0.015</td>
<td>0.022</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Power</td>
<td>0.003</td>
<td>0.001</td>
<td>0.063</td>
<td></td>
</tr>
<tr>
<td>Parking</td>
<td>0.034</td>
<td>0.034</td>
<td>0.299</td>
<td></td>
</tr>
<tr>
<td>Log of km per year (000s)</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Km per year at night (%)</td>
<td>-0.008</td>
<td>-0.005</td>
<td>0.009</td>
<td>0.017</td>
</tr>
<tr>
<td>Km per year at night (%)^2</td>
<td>0.0002</td>
<td>0.0001</td>
<td>0.0002</td>
<td>0.033</td>
</tr>
<tr>
<td>Km per year over speed Limit (%)</td>
<td>0.015</td>
<td>0.014</td>
<td>0.019</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Km per year over speed Limit (%)^2</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Urban km per year (%)</td>
<td>0.029</td>
<td>0.031</td>
<td>0.028</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>AIC</strong></td>
<td>29,631.281</td>
<td>30,624.100</td>
<td>29,809.179</td>
<td>29,658.447</td>
</tr>
<tr>
<td><strong>BIC</strong></td>
<td>29,736.934</td>
<td>30,689.117</td>
<td>29,857.942</td>
<td>29,707.210</td>
</tr>
<tr>
<td><strong>LogL</strong></td>
<td>-13,742.650</td>
<td>-14,244.060</td>
<td>-13,838.600</td>
<td>-13,763.230</td>
</tr>
<tr>
<td><strong>Chi-2</strong></td>
<td>1,357.220</td>
<td>&lt;0.001</td>
<td>1,165.320</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Concordant predictions of all models (in percentages).

<table>
<thead>
<tr>
<th>model results</th>
<th>All variables</th>
<th>Non-telematics</th>
<th>Telematics</th>
<th>Telematics with offsets</th>
</tr>
</thead>
<tbody>
<tr>
<td>All types of claims</td>
<td>62.28</td>
<td>55.91</td>
<td>61.34</td>
<td>62.10</td>
</tr>
<tr>
<td>Poisson model results with offsets (Log of Km per year in thousands). All types of claims</td>
<td>62.15</td>
<td>58.60</td>
<td>61.18</td>
<td>62.05</td>
</tr>
<tr>
<td>Claims where the policyholder is guilty</td>
<td>62.70</td>
<td>57.72</td>
<td>61.13</td>
<td>62.65</td>
</tr>
<tr>
<td>Poisson model results with offsets (Log of Km per year in thousands). Claims where the policyholder is guilty</td>
<td>62.38</td>
<td>58.96</td>
<td>60.89</td>
<td>62.43</td>
</tr>
</tbody>
</table>
Two step correction
Table IV. Zero-inflated Poisson model with offsets (Log of km per year in 000s). All types of claims.

<table>
<thead>
<tr>
<th></th>
<th>All variables</th>
<th>(Only significant)</th>
<th>Non-telematics</th>
<th>Telematics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>(p-value)</td>
<td>Coefficient</td>
<td>(p-value)</td>
</tr>
<tr>
<td>Poisson part</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-2.148</td>
<td>0.045</td>
<td>-3.396</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Age</td>
<td>-0.094</td>
<td>0.232</td>
<td>-0.123</td>
<td>0.121</td>
</tr>
<tr>
<td>Age$^2$</td>
<td>0.002</td>
<td>0.221</td>
<td>0.002</td>
<td>0.131</td>
</tr>
<tr>
<td>Male</td>
<td>-0.068</td>
<td>0.029</td>
<td>0.011</td>
<td>0.719</td>
</tr>
<tr>
<td>Age Driving Licence</td>
<td>-0.059</td>
<td>&lt;.001</td>
<td>-0.056</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Vehicle Age</td>
<td>0.014</td>
<td>&lt;.001</td>
<td>0.014</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Power</td>
<td>0.003</td>
<td>&lt;.001</td>
<td>0.003</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Parking</td>
<td>0.029</td>
<td>0.420</td>
<td>0.032</td>
<td>0.381</td>
</tr>
<tr>
<td>Log of km per year</td>
<td>1.000</td>
<td>--</td>
<td>1.000</td>
<td>--</td>
</tr>
<tr>
<td>(thousands) - offset</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Km per year at night (%)</td>
<td>-0.004</td>
<td>0.312</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Km per year at night (%)$^2$</td>
<td>0.0001</td>
<td>0.467</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Km per year over speed</td>
<td>0.019</td>
<td>0.001</td>
<td>0.019</td>
<td>0.001</td>
</tr>
<tr>
<td>limit (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Km per year over speed</td>
<td>-0.001</td>
<td>0.001</td>
<td>-0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>limit (%)$^2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban km per year (%)</td>
<td>0.026</td>
<td>&lt;.001</td>
<td>0.026</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Zero-inflation part</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept (Logit)</td>
<td>-0.847</td>
<td>&lt;.001</td>
<td>-0.857</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Log of km per year</td>
<td>0.404</td>
<td>&lt;.001</td>
<td>0.410</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>(thousands) (Logit)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>28.877.112</td>
<td></td>
<td>28.870.556</td>
<td></td>
</tr>
<tr>
<td>BIC</td>
<td>28.999.019</td>
<td></td>
<td>28.951.828</td>
<td></td>
</tr>
</tbody>
</table>
Concordant predictions of all models (in percentages).

<table>
<thead>
<tr>
<th>Zero Poisson model results with offsets (Log of Km per year in thousands). All types of claims</th>
<th>All variables</th>
<th>Non-telematics</th>
<th>Telematics</th>
<th>Telematics with offsets</th>
</tr>
</thead>
<tbody>
<tr>
<td>62.36</td>
<td>59.10</td>
<td>61.39</td>
<td>62.20</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Poisson model results with offsets (Log of Km per year in thousands). All types of claims</th>
<th>62.15</th>
<th>58.60</th>
<th>61.18</th>
<th>62.05</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Zero Poisson model results with offsets (Log of Km per year in thousands). Claims where the policyholder is at fault</th>
<th>62.71</th>
<th>59.85</th>
<th>61.17</th>
<th>62.77</th>
</tr>
</thead>
</table>

| Poisson model results with offsets (Log of Km per year in thousands). Claims where the policyholder is at fault | 62.38         | 58.96          | 60.89      | 62.43                  |
Changing driving habits: speed reduction
Cost of claims with telematics information
Conditional quantile as risk predictor
4 Going forward to optimal pricing
### Summary

- **Trip**
  - Linear models
  - Regression with categorical dependent variables
  - Regression with count-dependent variables
  - Generalized linear models
  - Frequency and severity models

- **Day**
  - Longitudinal and panel data models
  - Linear mixed models
  - Credibility and regression modeling
  - Fat-tailed regression models
  - Spatial modeling
  - Unsupervised learning

- **Week**
  - Bayesian regression models
  - Generalized additive models and nonparametric regression
  - Non-linear mixed models
  - Claims triangles/loss reserves
  - Survival models
  - Transition modeling

... and then correct premium
In dependent modelling claims, lapse and usage are all interconnected.
Pricing and Personalization

- Usage
  - Competition between insurers
  - Road safety, responsible driving
- Risk
  - Lapse
  - Claims
  - Company profit
  - Insurance market
Innovations create the demand for new insurance products for which there is no historical information and so, no mathematical way of measuring the risk of an accident.
### Challenge

The adaptation to digital innovations in the insurance companies themselves

1) Central role of data chief officer (CDO)

2) Promote CEOs cross-sectional vision of data analytics

3) Let data speak, Data-speak language is more than a number. **Analytics should express conclusions in sentences**, analysts should find the meaning to formulas, algorithms, figures and digits.
What have we learned?

1) The statistics on **driving style** are much more informative than the traditional rating factors

2) The level of **personalization** and the role of insurance changes

3) **Insurance** is reinvented in order to protect people and prevent accidents.

What comes ahead?

Insurance as a utility for protection, not only for compensation

Insurance pools

Autonomous/assisted driving. Joint ventures insurers-manufacturers
What have we learned?

1) The statistics on driving style are much more informative than the traditional rating factors.

2) The level of personalization and the role of insurance changes.

3) Insurance is reinvented in order to protect people and prevent accidents.

What comes ahead?

Insurance as a utility for protection, not only for compensation.

Insurance pools.

Autonomous/assisted driving. Joint ventures insurers-manufacturers.
See our work in progress: www.ub.edu/riskcenter
Driving data for automobile insurance: will telematics change ratemaking?

Montserrat Guillén

University of Barcelona
mguillen@ub.edu
www.ub.edu/riskcenter

SAV, Lucerne, Friday 30 August 2019