Driving data for automobile insurance: will telematics change ratemaking?

Montserrat Guillén

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

Riskcenter
UNIVERSITAT DE BARCELONA

SAV, Lucerne, Friday 30 August 2019

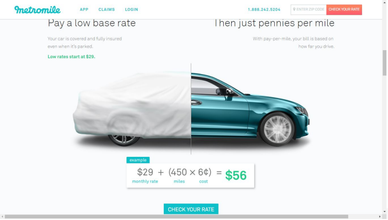
1 Introduction

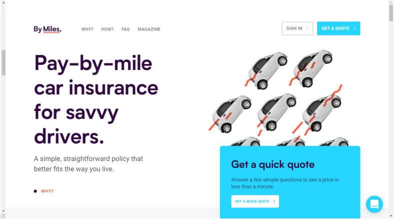
Sample Trip Summary Data – One Day

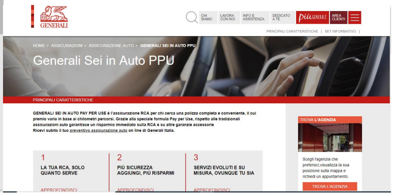
State Date	Start Time	Motorway Yards	Urban Yards	Other Yards	Motorway Seconds	Urban Seconds	Other Seconds	Total Speeding Yards	Total Speeding Seconds
3/3/2012	12:12:00	-	31	-	-	13,713	-	-	-
3/3/2012	14:17:11	-	3,355	-	-	7,934	-	-	-
3/3/2012	14:34:03	39,566	39,010	69,042	1,328	1,922	2,864	1,379	38
3/3/2012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

Source: Jim Janavich ideas.returnonintelligence.com

Companies selling motor insurance based on telematics







Riskcenter | Universitat de Barcelona

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

Riskcenter UNIVERSITAT DE BARCELONA

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

- 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 **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 **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**.

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)

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)

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)

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?

Riskcenter UNIVERSITAT DE BARCELONA

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:

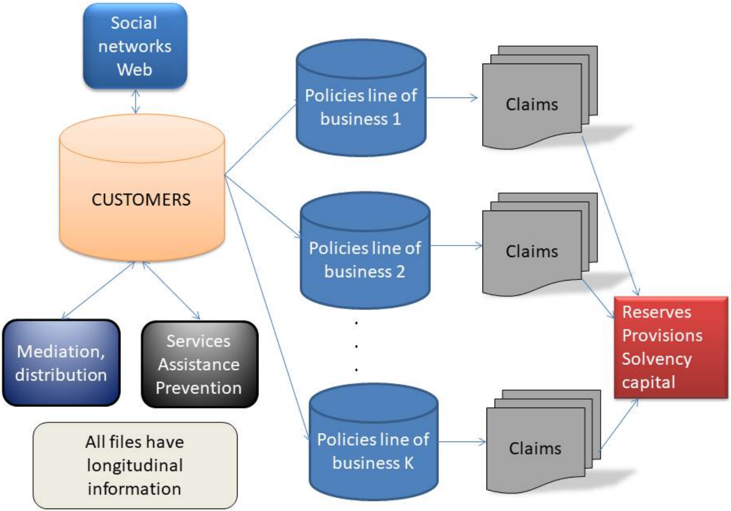
- Propose a method to **update premiums** regularly with telematics data. We create the basis for real-time pricing (not necessary), and real-time prevention.
- Show that the **price per mile depends on driving habits** and price should not be proportional to distance driven. A zero claim is relatively 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.

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.



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

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

  The unkown parameters to be estimated are $(\beta_0, \dots, \beta_k)$.
- 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.

Riskcenter UNIVERSITAT DE BARCELONA

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 ith 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 λ_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 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 ith 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 λ_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 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 ith 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 λ_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 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 ith 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 λ_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 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.

Transportation
https://doi.org/10.1007/s11116-018-9890-7

Improving automobile insurance ratemaking using telematics: incorporating mileage and driver behaviour data

Mercedes Ayuso[1] · **Montserrat Guillen**[1] · **Jens Perch Nielsen**[2]

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:

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

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:

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

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:

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

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:

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

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)

Risk Analysis

The use of telematics devices to improve automobile insurance rates

DOI:10.1111/risa.13172

Publication status

Article accepted on 9 July, 2018

Guillen, M. et al (2018)

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 Di indicates distance travelled, then:

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

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

$$P(Y_i = 0) = p_i + (1 - p_i)P(Y^* = 0) \tag{4}$$

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

$$p_i = \frac{exp(\alpha_0 + \alpha_1 ln(D_i))}{1 + exp(\alpha_0 + \alpha_1 ln(D_i)).} \quad (5)$$

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

$E(Y_i^*|x_i, T_i) = D_i exp(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}) = D_i \lambda_i = exp(ln(D_i))\lambda_i = exp(T_i)\lambda_i$, where $T_i = ln(D_i)$. The expectation of the Poisson part is:

$$(1 - p_i)E(Y_i^*|x_i, T_i) = \frac{1}{1 + exp(\alpha_0 + \alpha_1 ln(D_i))} D_i \lambda_i = D_i^* \lambda_i \quad (6)$$

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.

Riskcenter UNIVERSITAT DE BARCELONA

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

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

$$p_i = \frac{exp(\alpha_0 + \alpha_1 ln(D_i))}{1 + exp(\alpha_0 + \alpha_1 ln(D_i))} = \frac{exp(\alpha_0 + \alpha_1 ln(D_i))}{1 + exp(\alpha_0 + \alpha_1 ln(D_i))} \frac{D_i}{D_i} =$$
$$exp(\alpha_0+\alpha_1 ln(D_i)) \frac{D_i}{1 + exp(\alpha_0 + \alpha_1 ln(D_i))} \frac{1}{D_i} = exp(\alpha_0+\alpha_1 ln(D_i)) \frac{D_i^*}{D_i} \quad (7)$$

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

North American Actuarial Journal, 0(0), 1–11, 2019

ISSN: 1092-0277 print / 2325-0453 online
DOI: 10.1080/10920277.2019.1627221

Routledge
Taylor & Francis Group

Check for updates

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

Montserrat Guillen,[1], Jens Perch Nielsen,[2] Ana M. Pérez-Marín,[3] and Valandis Elpidorou[4]
[1]*Department of Econometrics, Riskcenter–IREA, Universitat de Barcelona, Barcelona, Spain*
[2]*Cass Business School, City, University of London, London, United Kingdom*
[3]*Department of Econometrics, Riskcenter–IREA, Universitat de Barcelona, Barcelona, Spain*
[4]*Arch Reinsurance Europe Underwriting dac Ireland, Dublin, Ireland*

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 a "near miss." A near miss—a name 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/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/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/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 $6 m/s^2$, 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/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?

3 Data and results


2009
MAPFRE,
la aseguradora global de confianza


YCAR
MAPFRE
NACE YCAR
EL SEGURO DE COCHE QUE
PIENSA EN LOS JÓVENES
Y EN SU BOLSILLO
MAPFRE


MAPFRE
2019
Manda tu pregunta
a Rafa Nadal
usando el hashtag
#MAPFREpreguntaRafa

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)

	Absolute frequency per driver		
Number of claims	All claims	Claims at fault	Claims not at fault
0	20,608	22,837	22,432
1	3,310	1,750	2,111
2	889	385	424
3	165	37	40
4	34	4	6
5	7	1	1
6	1	0	0

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)

	All Sample n = 25,014	Drivers with no claims n = 20,608 (82.4%)	Drivers with claims n = 4,406 (17.6%)
Mean	7.16	6.99	7.96
1st Quartile	4.14	4.00	4.87
Median	6.46	6.28	7.22
3rd Quartile	9.40	9.22	10.30
Standard Deviation	4.19	4.14	4.35

Table 2 Descriptive statistics by claims (quantitative variables)

	All sample N=25,014		Drivers with no claims N=20,608 (82.4%)		Drivers with claims N=4406 (17.6%)	
	Mean	SD	Mean	SD	Mean	SD
Age	27.57	3.09	27.65	3.09	27.18	3.10
Age driving licence	7.17	3.05	7.27	3.07	6.73	2.94
Vehicle age	8.75	4.17	8.76	4.19	8.69	4.11
Power	97.22	27.77	96.98	27.83	98.36	27.46
Km per year (000s)	7.16	4.19	6.99	4.14	7.96	4.35
Km per year at night (%)	6.91	6.35	6.85	6.32	7.16	6.49
Km per year over speed limit (%)	6.33	6.83	6.28	6.87	6.60	6.59
Urban km per year (%)	25.87	14.36	25.51	14.31	27.56	14.47

Table 3 Descriptive statistics by claims (categorical variables)

	All sample N=25,014		Drivers with no claims N=20,608 (82.4%)		Drivers with claims N=4406 (17.6%)	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
Gender						
Men	12,235	48.91	10,018	48.61	2217	50.32
Women	12,779	51.09	10,590	51.39	2189	49.68
Parking						
Yes	19,356	77.38	15,912	77.21	3444	78.17
No	5658	22.62	4696	22.79	962	21.83

Poisson model results. All types of claims.

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

	All variables		Non-telematics		Telematics		Telematics with offsets (Log of prediction of Non-telematics model - Column 2)	
	Coefficient	(p-value)	Coefficient	(p-value)	Coefficient	(p-value)	Coefficient	(p-value)
Intercept	-2.193	0.024	-0.472	0.625	-4.219	<.0001	-0.731	<.0001
Age	-0.145	0.043	-0.200	0.005				
Age^2	0.003	0.040	0.004	0.005				
Male	-0.086	0.002	-0.049	0.076				
Age Driving License	-0.061	<.0001	-0.076	<.0001				
Vehicle Age	0.015	<.0001	0.022	<.0001				
Power	0.003	<.0001	0.001	0.063				
Parking	0.034	0.292	0.034	0.299				
Log of km per year (000s)	1.000	--	1.000	--	1.000	--	1.000	--
Km per year at night (%)	-0.008	0.051			-0.005	0.161	-0.009	0.017
Km per year at night (%)^2	0.0002	0.062			0.0001	0.193	0.0002	0.033
Km per year over speed Limit (%)	0.015	0.004			0.014	0.006	0.019	<.001
Km per year over speed Limit (%)^2	-0.001	0.001			-0.001	0.003	-0.001	<.001
Urban km per year (%)	0.029	<.0001			0.031	<.0001	0.028	<.0001
AIC	29,631.281		30,624.100		29,809.179		29,658.447	
BIC	29,736.934		30,689.117		29,857.942		29,707.210	
LogL	-13,742.650		-14,244.060		-13,838.600		-13,763.230	
Chi-2	1,357.220	<0.001	354.400	<0.001	1,165.320	<0.001	1,316.060	<0.001

Riskcenter UNIVERSITAT DE BARCELONA

Concordant predictions of all models (in percentages).

	All variables	Non-telematics	Telematics	Telematics with offsets
Poisson model results. All types of claims	62.28	55.91	61.34	62.10
Poisson model results with offsets (Log of Km per year in thousands). All types of claims	62.15	58.60	61.18	62.05
Poisson model results. Claims where the policyholder is guilty	62.70	57.72	61.13	62.65
Poisson model results with offsets (Log of Km per year in thousands). Claims where the policyholder is guilty	62.38	58.96	60.89	62.43

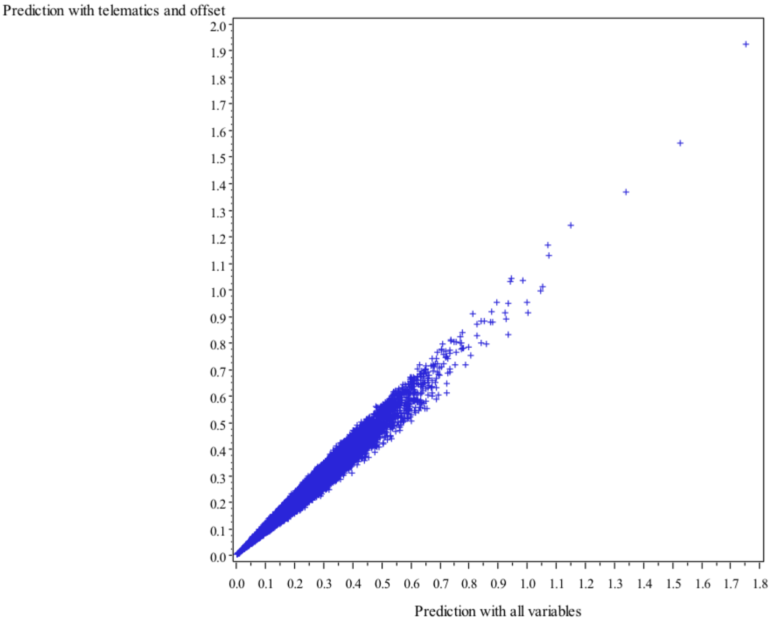

Prediction with telematics and offset
2.0
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8
Prediction with all variables

Table IV. Zero-inflated Poisson model with offsets (Log of km per year in 000s). All types of claims.

	All variables		(Only significant)		Non-telematics		Telematics	
	Coefficient	(p-value)	Coefficient	(p-value)	Coefficient	(p-value)	Coefficient	(p-value)
Poisson part								
Intercept	-2.148	0.045	-3.396	<.001	-0.829	0.440	-3.461	<.001
Age	-0.094	0.232			-0.123	0.121		
Age^2	0.002	0.221			0.002	0.131		
Male	-0.068	0.029	-0.074	0.017	-0.011	0.719		
Age Driving Licence	-0.059	<.001	-0.056	<.001	-0.067	<.001		
Vehicle Age	0.014	<.001	0.014	<.001	0.017	<.001		
Power	0.003	<.001	0.003	<.001	0.001	0.017		
Parking	0.029	0.420			0.032	0.381		
Log of km per year (thousands) - offset	1.000	--	1.000	--	1.000	--	1.000	--
Km per year at night (%)	-0.004	0.312					-0.001	0.771
Km per year at night $(\%)^2$	0.0001	0.467					0.000	0.931
Km per year over speed limit (%)	0.019	0.001	0.019	0.001			0.018	0.001
Km per year over speed limit $(\%)^2$	-0.001	0.001	-0.001	0.001			-0.001	0.003
Urban km per year (%)	0.026	<.001	0.026	<.001			0.027	<.001
Zero-inflation part								
Intercept (Logit)	-0.847	<.001	-0.857	<.001	-1.639	<.001	-0.795	<.001
Log of km per year (thousands) (Logit)	0.404	<.001	0.410	<.001	0.824	<.001	0.406	<.001
AIC	28,877.112		28,870.556		29,427.423		29,005.172	
BIC	28,999.019		28,951.828		29,508.694		29,070.189	

Riskcenter UNIVERSITAT DE BARCELONA

Concordant predictions of all models (in percentages).

	All variables	Non-telematics	Telematics	Telematics with offsets
Zero Poisson model results with offsets (Log of Km per year in thousands). All types of claims	62.36	59.10	61.39	62.20
Poisson model results with offsets (Log ok Km per year in thousands). All types of claims	62.15	58.60	61.18	62.05
Zero Poisson model results with offsets (Log of Km per year in thousands). Claims where the policyholder is at fault	62.71	59.85	61.17	62.77
Poisson model results with offsets (Log ok 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

• Linear models	• Longitudinal and panel data models	• Bayesian regression models
• Regression with categorical dependent variables	• Linear mixed models	• Generalized additive models and nonparametric regression
• Regression with count-dependent variables	• Credibility and regression modeling	• Non-linear mixed models
• Generalized linear models	• Fat-tailed regression models	• Claims triangles/loss reserves
• Frequency and severity models	• Spatial modeling	• Survival models
	• Unsupervised learning	• Transition modeling

… and then correct premium

In dependent modelling claims, lapse and usage are all interconnected

Pricing and Personalization

Riskcenter UNIVERSITAT DE BARCELONA

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

Denuit, M., Guillen, M. and Trufin, J. (2019) “Multivariate credibility modeling for usage-based motor insurance pricing with behavioural data” **Annals of Actuarial Science**, 13, 2, 378-399.

Pérez-Marín, A.M. and Guillen, M. (2019) “The transition towards semi-autonomous vehicle insurance: the contribution of usage-based data ”, **Accident Analysis and Prevention**, 123, 99-106.

Pérez-Marin, A.M., Ayuso M.M. and Guillen, M. (2019) “Do young insured drivers slow down after suffering an accident?”, **Transportation Research Part F: Psychology and Behaviour**, 62, 690-699.

Guillen, M., Nielsen, J.P., Ayuso, M. and Pérez-Marín, A.M. (2019) “The use of telematics devices to improve automobile insurance rates”, **Risk Analysis**, 39, 3, 662-672.

Ayuso, M., Guillen, M. and Nielsen, J.P. (2019) “Improving automobile insurance ratemaking using telematics: incorporating mileage and driver behaviour data”, **Transportation**, 46(3), 735-752. .

Boucher, J.P., Coté, S. and Guillen, M. (2018) “Exposure as duration and distance in telematics motor insurance using generalized additive models”, **Risks**, 5(4), 54.

Ayuso, M., Guillen, M., Pérez-Marín, A.M. (2016a) “Telematics and gender discrimination: some usage-based evidence on whether men’s risk of accident differs from women’s”, **Risks**, 2016, 4, 10.

Ayuso, M., Guillen, M., and Pérez-Marín, A.M. (2016b) “Using GPS data to analyse the distance travelled to the first accident at fault in pay-as-you drive insurance”, **Transportation Research Part C** 68, 160-167.

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