Dynamically Updating Motor Insurance Prices with Telematics Collected Driving Behavior Data

Katrien Antonio

LRisk - KU Leuven and
ASE - University of Amsterdam

EAA e-Conference on
Data Science & Data Ethics

12 May 2022

Presenter + Authors

Katrien Antonio

Roel Henckaerts

Paper will appear in Insurance: Mathematics and Economics.

Recent work on insurance pricing analytics

[Henckaerts et al., 2018]

SAJ

[Henckaerts et al., 2021]

NAAJ

[Henckaerts et al., 2022]

Expert Syst. Appl.

Page 2

[Henckaerts & Antonio, 2022]

IME

distRforest

github/henckr/distRforest

maidrr

github/henckr/maidrr

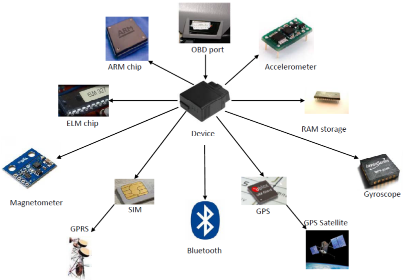
- Denote for policyholder i in a given policy period:
  - e_i: exposure-to-risk
  - N_i: number of claims filed during the exposure period
  - L_i: total loss amount reported during the exposure period.
- The pure premium π_i:

$$\pi_i = \mathbb{E}\left[\frac{L_i}{e_i}\right] \overset{indep.}{=} \mathbb{E}\left[\frac{N_i}{e_i}\right] \times \mathbb{E}\left[\frac{L_i}{N_i} \mid N_i > 0\right] = \underbrace{Freq_i}_{\textbf{frequency}} \times \underbrace{\text{Sev}_i}_{\textbf{severity}}$$

- Build f(risk factors) to predict frequency and severity, respectively.

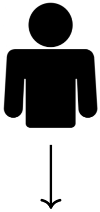
Telematics insurance

Products: usage-based insurance (UBI)
pay-as-you-drive (PAYD)
pay-how-you-drive (PHYD)



- **Telematics** is the integrated use of **tele**communications and infor**matics**.
- Black-box device is installed in the vehicle.
- **Real driving behavior** is monitored.
- Very often targets **young drivers**.

Risk factors for motor insurance pricing

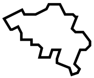


Static, demographic data

License age

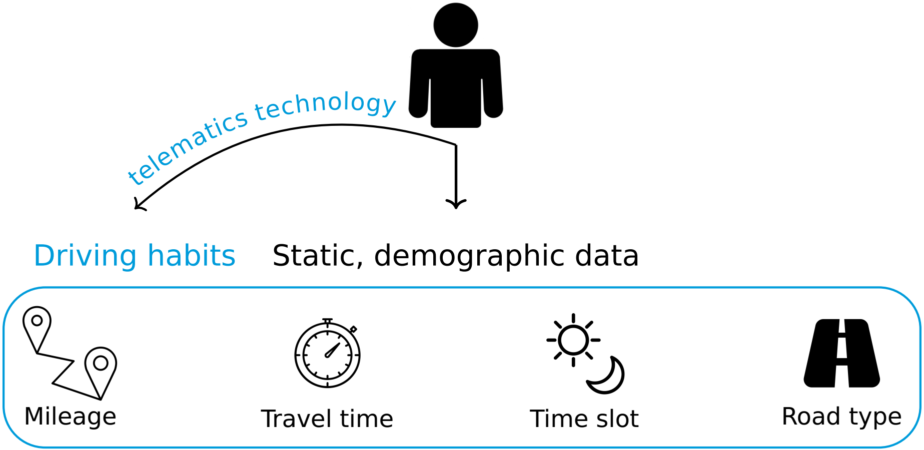
Car make/model

Type of fuel

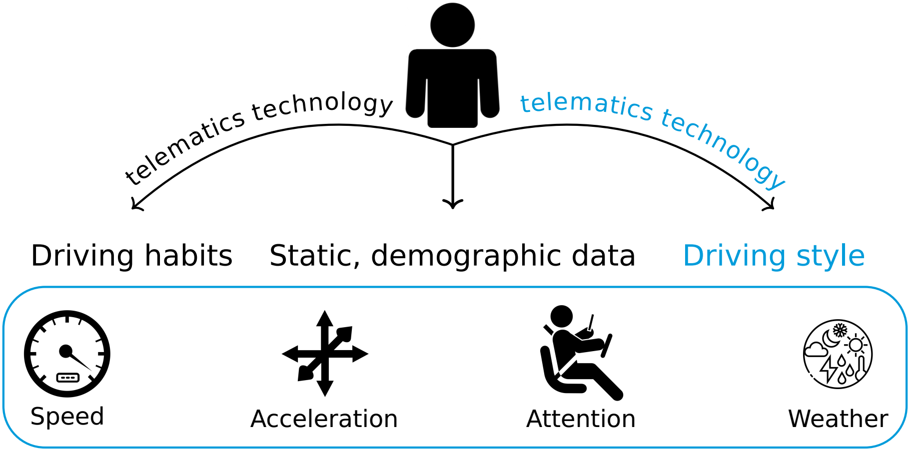


Postal code

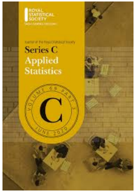
Risk factors for motor insurance pricing



Risk factors for motor insurance pricing

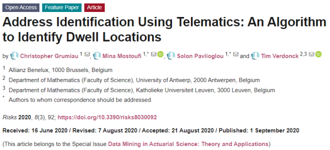


Insurance analytics literature on telematics



- Verbelen, Antonio & Claeskens (2018, JRSS C):
  - claim frequency models with classic, static features and driving habit information
  - compositional data and their use in GAMs.
- Wüthrich (2017, EAJ), Gao & Wüthrich (2018, EAJ), Gao et al. (2019, SAJ) and more papers:
  - the construction of $v - a$ **heatmaps from GPS signals**
  - feature-engineering on these heatmaps
  - use of these features in claim frequency models.

Insurance analytics literature on telematics


Open Access | Feature Paper | Article

Address Identification Using Telematics: An Algorithm to Identify Dwell Locations

by Christopher Grumiau [1], Mina Mostoufi [1,*], Solon Pavlioglou [1,*] and Tim Verdonck [2,3]

[1] Allianz Benelux, 1000 Brussels, Belgium
[2] Department of Mathematics (Faculty of Science), University of Antwerp, 2000 Antwerpen, Belgium
[3] Department of Mathematics (Faculty of Science), Katholieke Universiteit Leuven, 3000 Leuven, Belgium
[*] Authors to whom correspondence should be addressed.

Risks 2020, 8(3), 92; https://doi.org/10.3390/risks8030092

Received: 16 June 2020 / Revised: 7 August 2020 / Accepted: 21 August 2020 / Published: 1 September 2020

(This article belongs to the Special Issue Data Mining in Actuarial Science: Theory and Applications)

- Denuit, Guillen & Trufin (2019, Annals of Actuarial Science) on ***Multivariate credibility modelling for usage-based motor insurance pricing with behavioural data***.

- Grumiau, Mostoufi, Pavlioglou & Verdonck (2020, Risks) on ***Address identification using telematics: an algorithm to identify dwell locations***.

- Banghee So, J.-P. Boucher & E. Valdez (2021, Risks) on **Synthetic dataset generation of driver telematics**.

Managerial insights, based on Carbone & Taub (2018) UBI insurance is not usage-based. Sorry, not sorry!

- In 2017, 14 million policies sent telematics data to insurers around the world.
- However, less than 9 percent of the global insurance telematics policies were characterized by **usage-based pricing**.
- **Use of driving data in pricing**:
  - use driving score at underwriting stage
  - propose tailored renewal price (with discounts, or discounts + surcharges)
  - usage-based, i.e. charge price for period of coverage based on how policyholder behaves during this period, and avoid **premium leakage**.

Research questions

Our focus in this talk:

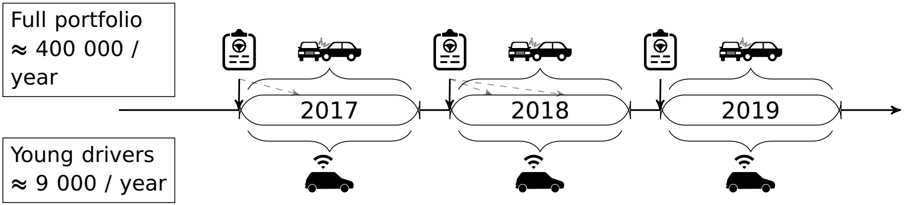
- How to use driving behavior (i.e. habits + style) to **update** a baseline **tariff** (with only self-reported characteristics)?
- What is the **added value** of telematics for pricing via risk classification?
- Managerial insights? Impact on retention rates, profit?

Focus on **frequency, severity and churn models** in the presence of static self-reported characteristics as well as telematics collected data.

Aim for an **explainable** updating mechanism.

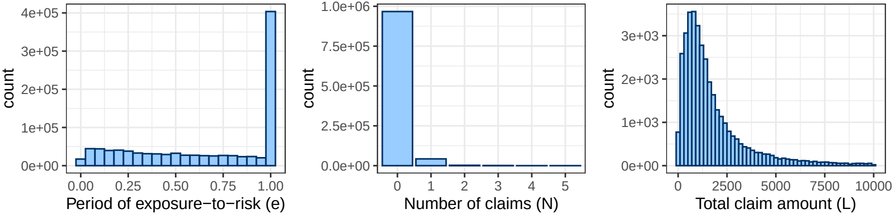
Data and methodology

Motor third party liability (MTPL) portfolio



- Policy information at the start of the policy period, subject to possible changes during the policy period (e.g., new vehicle).
- Claims reported (68 196 in total).
- Driving behavior (only for drivers < 26 years at underwriting time, 308M kilometers driven in total).

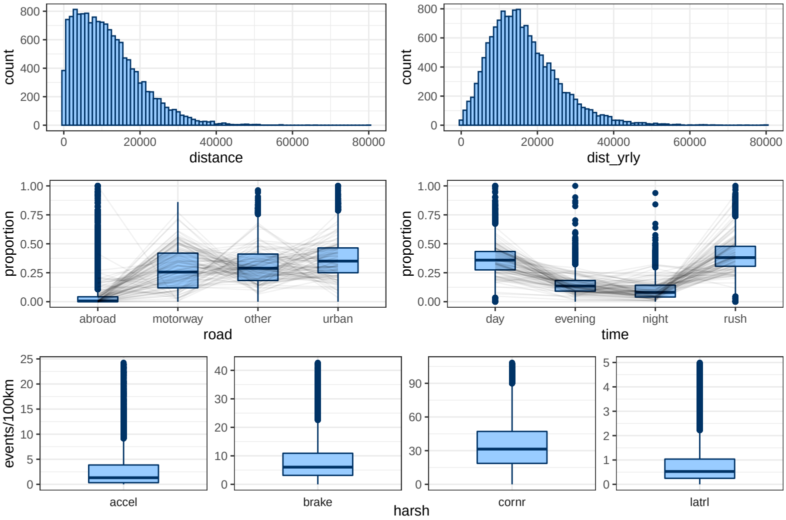
Claims and policy information for the complete portfolio



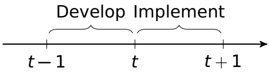
Policy information with **self-reported** risk characteristics:

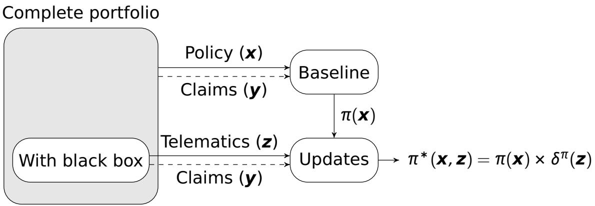
- **driver**: age, experience, additional young drivers, etc.
- **payments**: frequency and SEPA indicator
- **geographical**: postal code and mosaic segment
- **vehicle**: age, weight, value, power, fuel, make, etc.

Driving behavior information for young drivers



Updating methodology





The idea:

- charge baseline tariff $\pi(\boldsymbol{x})$ at t
- ex post, multiplicative update $\delta^\pi(\boldsymbol{z})$ at $t+1$, based on driving data in $[t, t+1]$.

Baseline pricing and churn models

Baseline model training

- Predictive models for the **complete portfolio** using traditional features $\boldsymbol{x}$
  - claim frequency *and* severity → **tariff**
  - customer churn prediction → client **retention** analysis.
- Stochastic gradient boosting (Friedman, 2002) with the following **assumptions**:

	Distribution	Prediction $f(\boldsymbol{x})$	Loss function $D(y, f(\boldsymbol{x}))$
Claim frequency	$N \sim$ Poisson	$\mathbb{E}(N \mid \boldsymbol{x}, e)$	$\frac{2}{n}\sum_{i=1}^{n}\left[y_i \ln\left\{\frac{y_i}{f_i}\right\} - \{y_i - f_i\}\right]$
Claim severity	$L/N \sim$ gamma	$\mathbb{E}(L/N \mid \boldsymbol{x})$	$\frac{2}{\sum_i N_i}\sum_{i=1}^{n} N_i\left[\frac{y_i - f_i}{f_i} - \ln\left\{\frac{y_i}{f_i}\right\}\right]$
Customer churn	$C \sim$ Bernoulli	$\mathbb{E}(C \mid \boldsymbol{x})$	$-\frac{1}{n}\sum_{i=1}^{n}\left[y_i \ln\{f_i\} + (1-y_i)\ln\{f_i\}\right]$

- Parameter **tuning**:

  H2O random grid search + 5-fold cross-validation (LeDell et al., 2020).

- Enforce the **balance property** by scaling predictions (for the young drivers):

  $$\sum_{i=1}^{n} f_i = \sum_{i=1}^{n} y_i$$

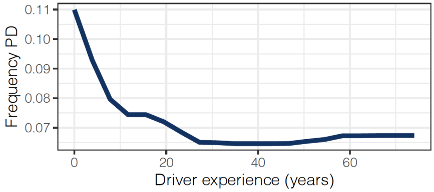
  automatically fulfilled for GLMs with canonical link

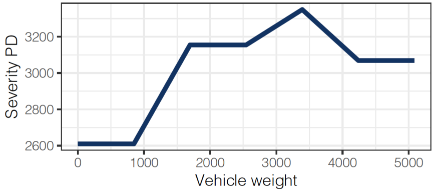
  see e.g. Wüthrich (2020).

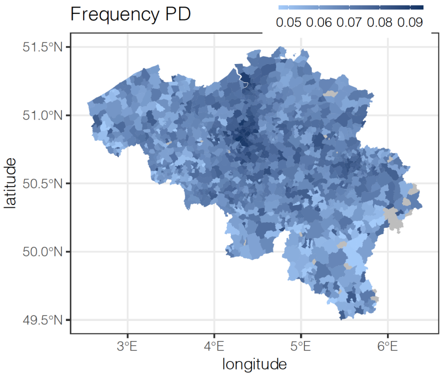
Insights in the optimal GBMs

	Claim frequency		Claim severity		Customer churn	
Rank	Feature	%	Feature	%	Feature	%
1	geo_postcode	34.72	veh_weight	23.21	paym_split	43.48
2	driv_experience	14.08	veh_make	21.37	geo_postcode	11.67
3	driv_seniority	8.52	geo_postcode	10.54	veh_age	9.85
4	veh_make	6.25	veh_segment	10.48	paym_sepa	9.44
5	geo_mosaic	5.85	geo_mosaic	6.59	driv_seniority	6.90
6	veh_fuel	5.09	driv_seniority	5.83	veh_make	3.43
7	veh_segment	4.66	veh_value	3.50	driv_experience	2.85
8	paym_split	3.91	veh_age	3.44	geo_mosaic	2.45
9	driv_add_younger26	3.29	driv_experience	2.98	driv_age	2.43
10	driv_age	2.75	driv_add_younger26	2.91	veh_use	1.99
$\sum$		89.12		90.86		94.48

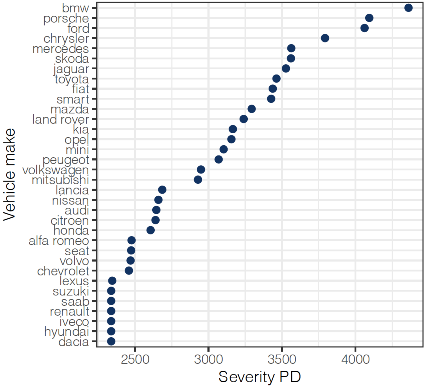
Insights in the optimal GBMs (cont.)

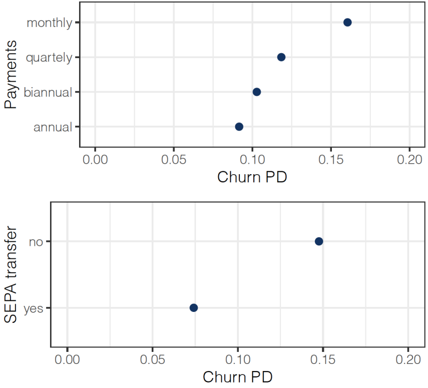






Insights in the optimal GBMs





Updating pricing with driving behavior

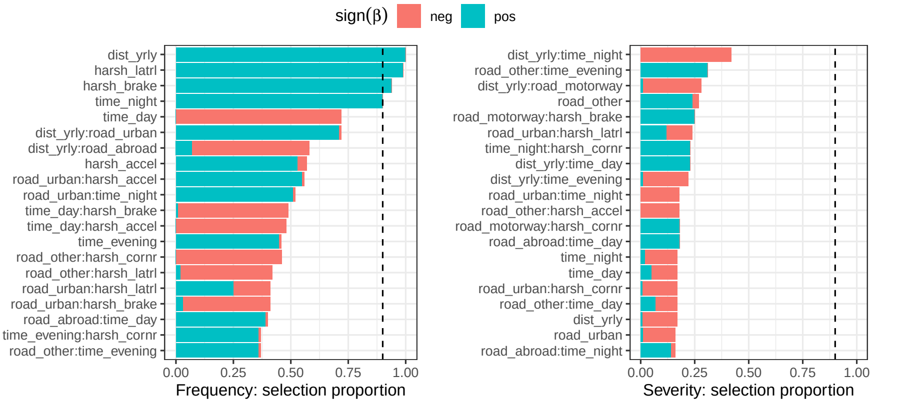
Update baseline premium with telematics

- Aim is to update premiums for the drivers with telematics features $\boldsymbol{z}$.
- Log-link GLM with the baseline prediction $\ln[f(\boldsymbol{x})]$ as an offset:

$$\begin{aligned}\ln[\mathbb{E}(y\,|\,\boldsymbol{x},\boldsymbol{z})] &= \ln[f(\boldsymbol{x})] + \beta_0 + \sum_{j=1}^{p} \beta_j z_j \\ \mathbb{E}(y\,|\,\boldsymbol{x},\boldsymbol{z}) &= f(\boldsymbol{x}) \times \exp(\beta_0) \times \prod_{j=1}^{p} \exp(\beta_j z_j).\end{aligned}$$

- Updated prediction is then **multiplicative**:
  - baseline GBM prediction $f(\boldsymbol{x})$ for a policyholder with risk characteristics $\boldsymbol{x}$
  - overall update factor $\exp(\beta_0)$ via the intercept
  - update $\exp(\beta_j z_j)$ from each telematics feature z_j.

Telematics feature selection with LASSO



- Let $\boldsymbol{z}^* \in \mathbb{R}^4$ denote {`dist_yrly`, `harsh_latrl`, `harsh_brake`, `time_night`}.
- Log-link Poisson GLM with offset for **claim frequency**:

$$\ln[\mathbb{E}(N \,|\, \boldsymbol{x}, \boldsymbol{z}^*)] = \ln[\mathbb{E}(N \,|\, \boldsymbol{x}, e)] + \beta_0 + \sum_{j=1}^{4} \beta_j \log(z_j^* + 1)$$

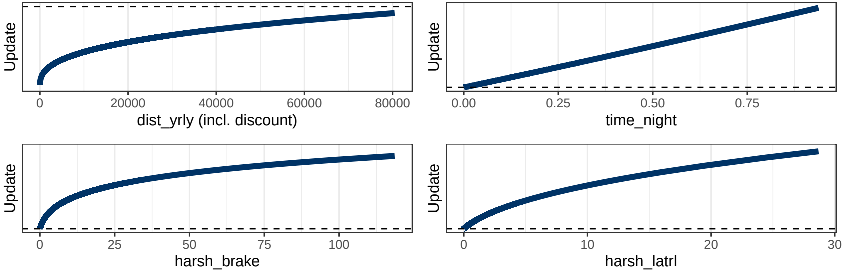
$$\mathbb{E}(N \,|\, \boldsymbol{x}, \boldsymbol{z}^*) = \mathbb{E}(N \,|\, \boldsymbol{x}, e) \times \exp(\beta_0) \times \prod_{j=1}^{4} (z_j^* + 1)^{\beta_j}.$$

Update price via the claim frequency component

Updated prediction is **multiplicative** in the following terms:

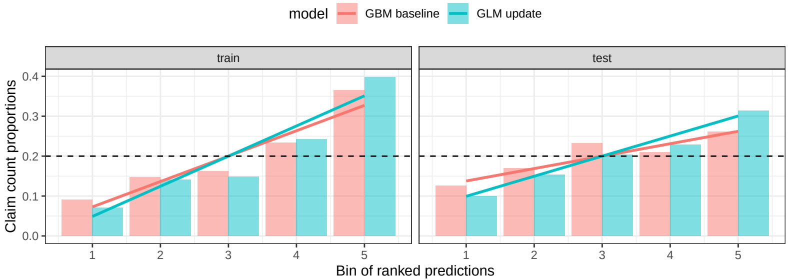
- baseline GBM prediction $\mathbb{E}(N|\boldsymbol{x}, e)$ for a policyholder with risk characteristics $\boldsymbol{x}$
- overall discount factor $\exp(\beta_0) \approx 2\%$
- update $(z_j^* + 1)^{\beta_j}$ from each telematics feature z_j^*.

Multiplicative update effects



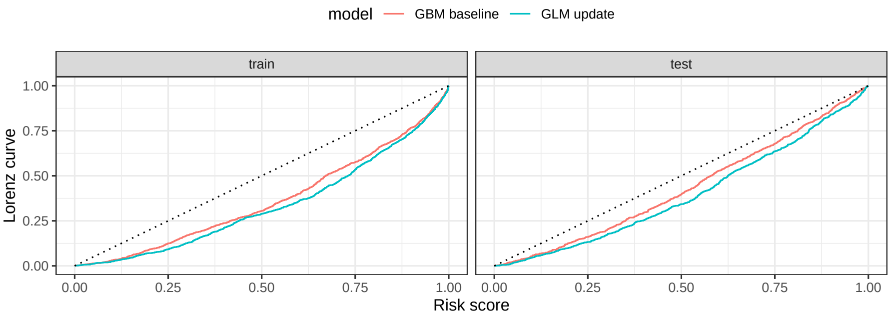
- Mileage + discount remains < 1.
- Penalty once night driving, harsh braking or lateral events are registered.
- Safe driving is the key to earn discounts!

Telematics improving risk classification



Here we consider:

- $r_i^m = F_n\left(f^m(\boldsymbol{x}_i, \boldsymbol{z}_i^*)\right)$ with $F_n(.)$ the ecdf
- $PC^m(s) = \frac{\sum_{i=1}^n N_i \mathbb{1}\{\frac{s-1}{5} < r_i^m \leq \frac{s}{5}\}}{\sum_{i=1}^n N_i}$ for $s \in \{1, \ldots, 5\}$.



Here we consider:

- $LC^m(s) = \frac{\sum_{i=1}^n N_i \, \mathbb{1}\{r_i^m \leq s\}}{\sum_{i=1}^n N_i}$ for $s \in [0, 1]$.

Managerial insights

Managerial insights

For the discussion of managerial insights, I refer to our paper:

- **adjust baseline churn** $\rho(\boldsymbol{x})$ to $\rho^*(\boldsymbol{x}, \delta^\pi) = \rho(\boldsymbol{x}) + \epsilon_p \cdot (\delta^\pi - 1) = \rho(\boldsymbol{x}) + \delta^\rho$, with ϵ_p the price elasticity
- study expected **profit and retention rate**

$$P = \frac{1}{n}\sum_{i=1}^{n}(1-(\rho_i+\delta_i^\rho))\cdot(\delta_i^\pi \pi_i - L_i) \qquad R = \frac{1}{n}\sum_{i=1}^{n} 1-(\rho_i+\delta_i^\rho)$$

- restrict penalties/discounts + redistribute ⇒ **fairness, solidarity, commercially appealing**

$$\delta_{lo}^\pi \leq \delta^\pi \leq \delta_{hi}^\pi$$

$$\sum_{i=1}^{n}(1-\rho_i)\cdot\pi_i = \sum_{i=1}^{n}(1-\rho_i)\cdot\alpha\cdot\delta_i^\pi\cdot\pi_i$$

Conclusions

Our paper puts focus on:

- a baseline pricing model with self-reported characteristics
- an explainable updating mechanism to incorporate driving behavioral information.

Added value of telematics for insurance pricing is studied from both a statistical and managerial perspective.

More information

For more information, please visit:

- draft of the paper, including complete set of references
- LRisk website, `www.lrisk.be`
- my homepage `https://katrienantonio.github.io`.

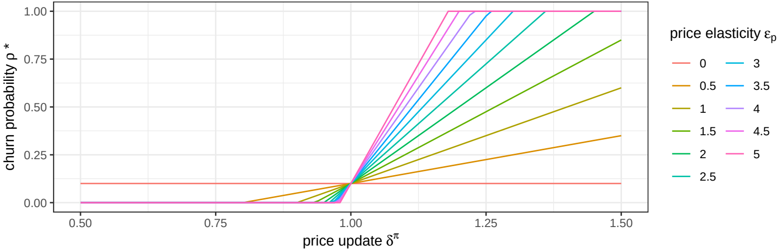
Special thanks to

- the organizers
- the companies and funding agencies supporting/having supported my research lab: Ageas, Argenta, Atlas Copco, CNP Assurances, FWO, KU Leuven internal funds.

Extra Sheets

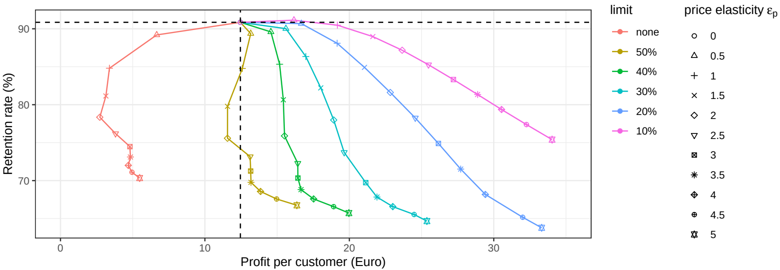
Churn updates

- Adjust baseline churn via price **update** δ^{π} and **elasticity** of demand ϵ_p.
- $\rho^*(\boldsymbol{x}, \delta^{\pi}) = \rho(\boldsymbol{x}) + \epsilon_p \cdot (\delta^{\pi} - 1)$ for $\rho(\boldsymbol{x}) = 0.1$ and $\epsilon_p \in [0, 5]$:



- $\rho^* = \rho$ when $\delta^{\pi} = \pi^*/\pi = 1$ (no price change).
- Linear increases/decreases ($\delta^{\pi} > 1$ / $\delta^{\pi} < 1$) with slope ϵ_p.

- Expected profits and retention rates under different scenarios:



- Stricter limits result in higher profits
- Profits increase with ϵ_p at the cost of lower retention
- No limit results in lower profits than baseline (driven by low premiums on average)

Constrained optimization for profit maximization

- Maximize expected profit P while retaining a minimum proportion of the portfolio R^*:

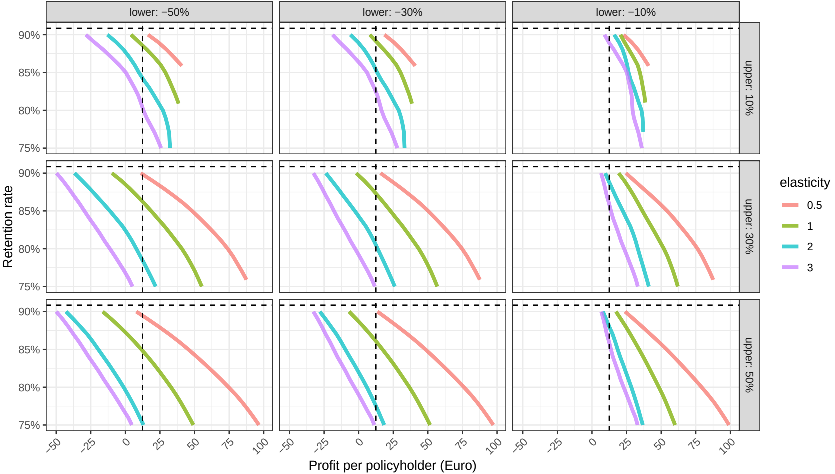
$$\max_{\alpha} P(\alpha) = \frac{1}{n} \sum_{i=1}^{n} (1 - (\rho_i + \delta_i^{\rho})) \cdot (\alpha \delta_i^{\pi} \pi_i - L_i)$$

$$\text{subject to } R(\alpha) = \frac{1}{n} \sum_{i=1}^{n} 1 - (\rho_i + \delta_i^{\rho}) \geq R^*$$

$$\delta_{lo}^{\pi} \leq \delta^{\pi} \leq \delta_{hi}^{\pi}.$$

- Implicit dependence of R on α as $\delta^{\rho} = \epsilon_p \cdot (\alpha \delta^{\pi} - 1)$.
- Efficient frontier by varying R^* over a range of values and maximizing $P(R^*)$ via α.

Efficient frontiers for $R^* \in [0.75, 0.9]$ under different scenarios



EAA e-Conference on
Data Science & Data Ethics

12 May 2022

Thank you very much for your attention

Contact

Katrien Antonio

LRisk - KU Leuven and
ASE - University of Amsterdam

katrien.antonio@kuleuven.be