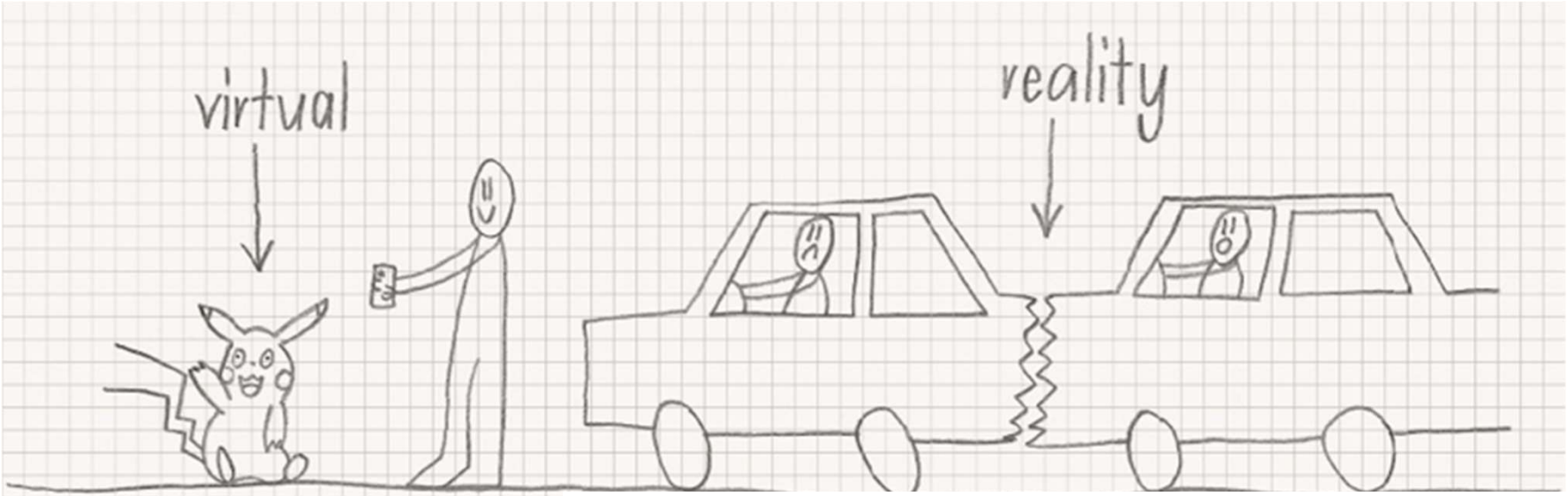
die Mobiliar



Explainable AI in Non-Life Insurance Pricing

Michael Mayer, Jan. 16, 2020 in Warsaw

Outline

1. Non-Life Insurance Pricing
2. The Role of XAI

Disclaimer

The opinions expressed in this presentation are not necessarily those of my employer. Please do not mistake my opinions for theirs.

Swiss Mobiliar at a Glance

- Founded in 1826
- >2 million customers*
- CHF 3.8 billion in premiums*
- CHF 443.5 million profit*
- Organised as a cooperative
- >5'000 employees
- About ten non-life pricing actuaries

1 CHF = 0.91 Euro = 3.88 PLN

*https://www.mobiliar.ch/die-mobiliar/ueber-uns

die Mobiliar

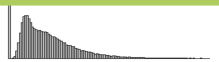
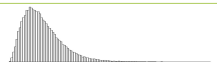
1. Non-Life Insurance Pricing

What is Non-Life Insurance Pricing?

"Non-life insurance pricing is the art of setting the price of an insurance policy, taking into consideration various properties of the insured object and the policy holder. The main source on which to base the decision is the insurance company's own historical data on policies and claims [...]. In a *tariff analysis*, the actuary uses this data to find a model which describes how the claim cost of an insurance policy depends on a number of explanatory variables."

- Preface, Ohlsson and Johansson (2010)

What does it mean from Statistical Perspective?

Response	Meaning	Distribution	Approach	Standard Metric	Sample Weights
Claim cost	Losses ($$$) / Exposure		Frequency * Severity; Tweedie-GLM	Tweedie deviance?	Exposure
Frequency	Claim count / Exposure		Poisson-like GLM with log-link	Poisson deviance	Exposure
Severity	Losses / Claim count		Gamma-GLM with log-link	Gamma deviance	Claim count

Exposure: E.g. the policy duration in years

$$Claim\ Cost = \frac{Losses}{Exposure} = \frac{Claim\ Count}{Exposure} \cdot \frac{Losses}{Claim\ Count} = Frequency \cdot Severity$$

Data structure for illustration

	Policy	Year	Exposure	Driver age	Claim count	Losses	Claim cost	Frequency	Severity
1	a	2017	0.25	23	0	0	0	0	NaN
2	a	2017	0.75	24	0	0	0	0	NaN
3	a	2018	0.25	24	1	5000	20000	4	5000
4	a	2018	0.75	25	0	0	0	0	NaN
5	b	2017	1.00	65	2	30000	30000	2	15000
6	b	2018	1.00	66	0	0	0	0	NaN

Only these rows are used in Severity model (other rows have sample weight 0)

Challenges?

Illustration with Real Data

Car Collisions with Large Animals

Annual figures for Switzerland (2015)

- Many large animals die on Swiss roads, among them 8'000 deers.
- Ca. 100 persons injured.
- Total vehicle damage: 25 Mio CHF
- Covered by partial damage coverage of motor insurance.

Source: https://de.wikipedia.org/wiki/Wildunfall

Annual figures for Swiss Mobiliar

- Ca. 2'000 animal collision claims
- Costs: many Mio CHF
- Roughly 500’000 exposure
- Low claim frequency of 2'000 / 500'000 = 0.4%

Model for claim frequency, taking into account individual risk factors?

GLM for Claim Frequency

Data

Car policies with partial damage coverage (2013 – 2018)

# Rows	Exposure *w*	Claim count	Frequency *y*
8'591'192	3'111'446	12'916	0.00415

Split

- 80% training, 10% validation, 10% reserve
- Grouped by policy number

Covariables

- Driver: place of living, age, gender, …
- Car: price, age, weight-to-power ratio, leased, …
- Policy: bonus protection, fully or partially comprehensive, year, …

Poisson-GLM with log-link

$$E(y) = \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m)$$

Estimates of $\boldsymbol{\beta}$ minimize Poisson deviance

$$2\sum w_i (y_i \log(y_i / \hat{y}_i) - (y_i - \hat{y}_i))$$

Results on training data

% deviance explained (training): 3.44%

Regressor	Estimates
log_catalogue_price	-0.0810
driver_age	-0.0242
max_7000_km (no)	0.5911
vehicle_age	-0.0225
town (no)	0.7786
…	…

2. The Role of XAI

Why still GLMs – and why XAI will be Important?

1. Important source of information (e.g. for tariff development),
2. Transparency (management, client, …)
3. Industry standard
4. Well supported by standard software
5. Model structure compatible with tariff formula

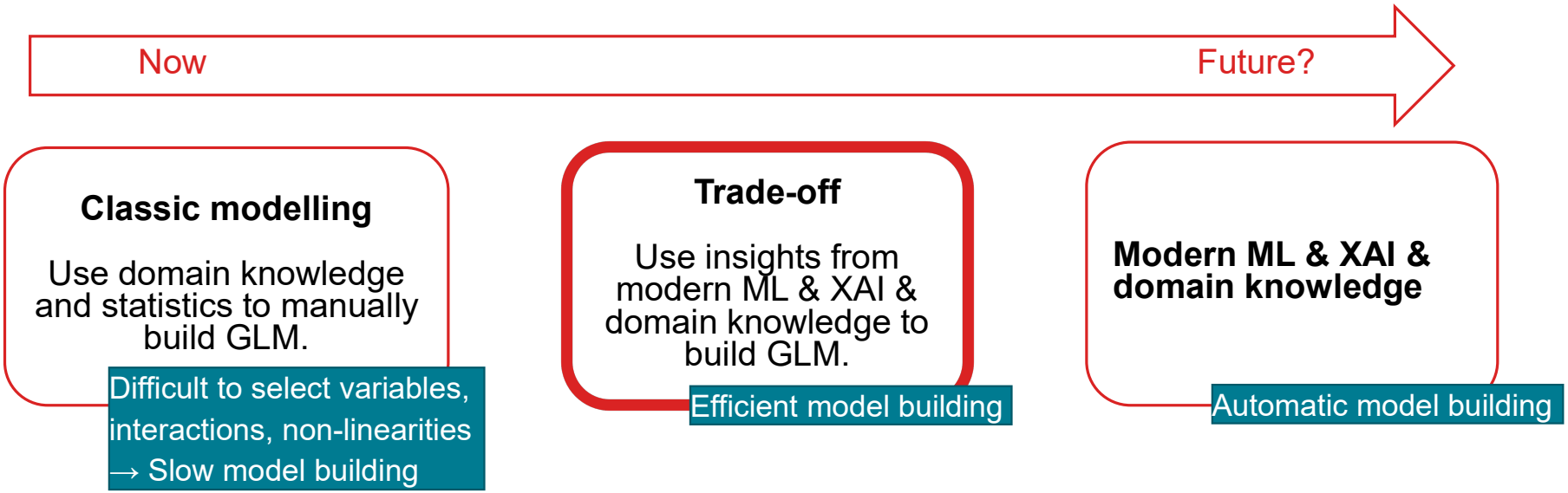
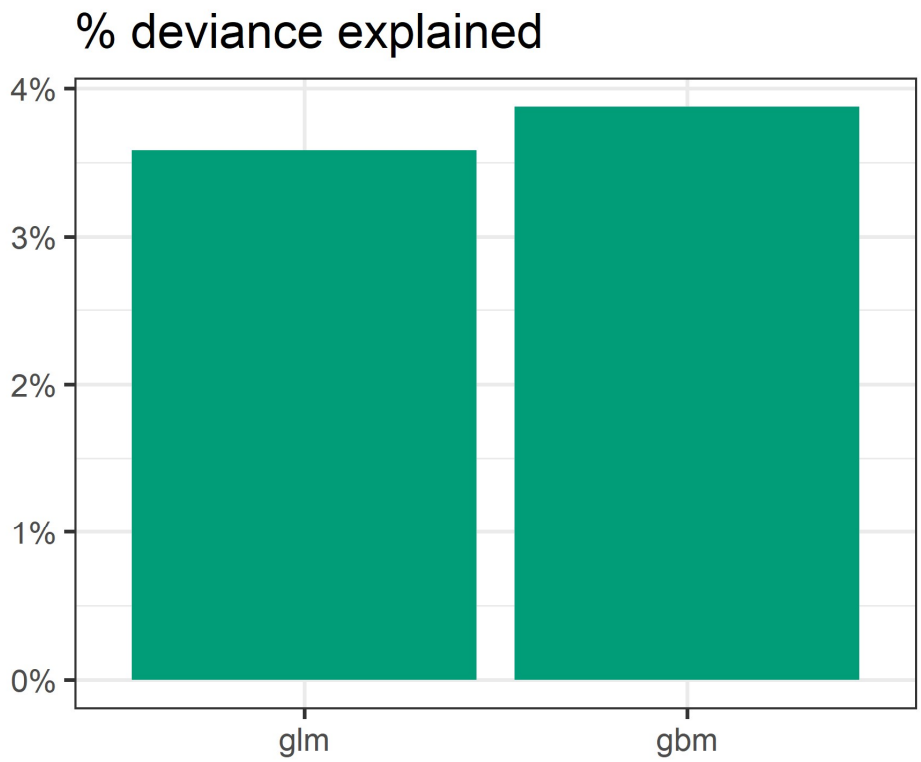


Illustration of Trade-Off

Let's Improve GLM by Insights from Tuned lightGBM Model

Performance on Validation Data?



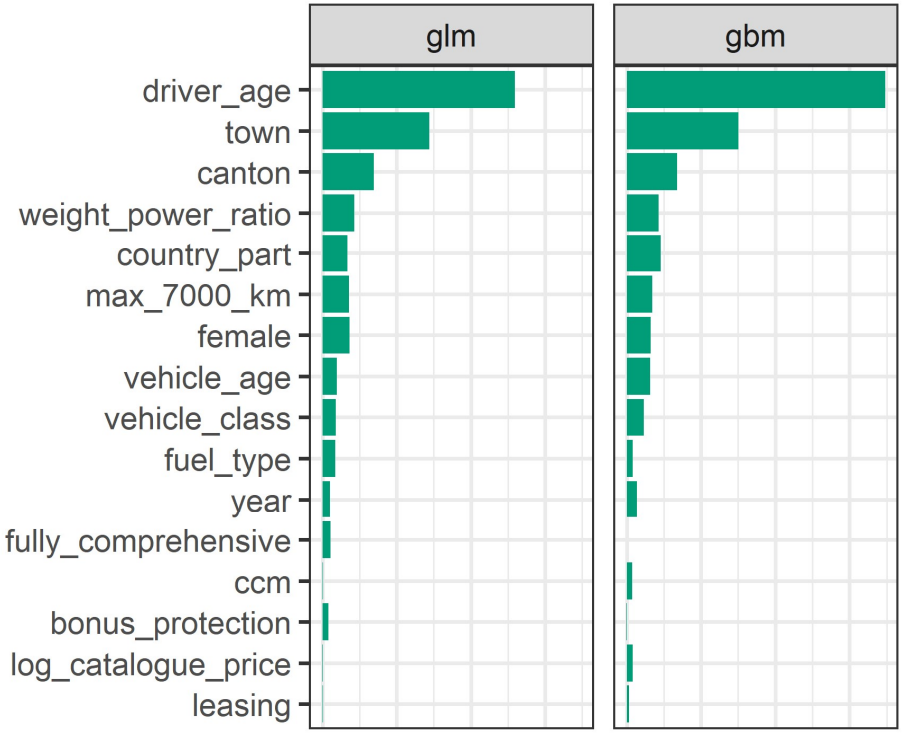
Both models use Poisson deviance loss

Insights

- Results make sense.
- There is room for improvement for GLM, but not too much.

Which Variables are most Important?

Permutation importance on validation data regarding Poisson deviance



Insights

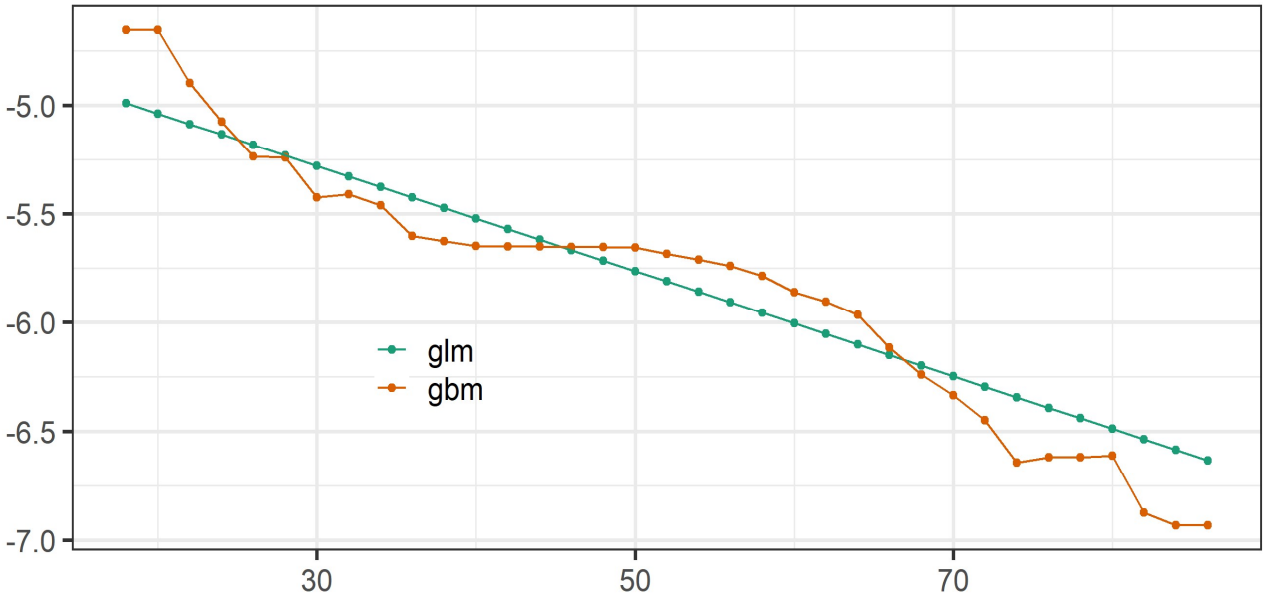
- Results make sense.
- Could remove
  - fully comprehensive (y/n),
  - bonus protection

  as not deemed to be relevant for animal collisions.

Do Main Effects make Sense? Do we need to adapt them?

Task

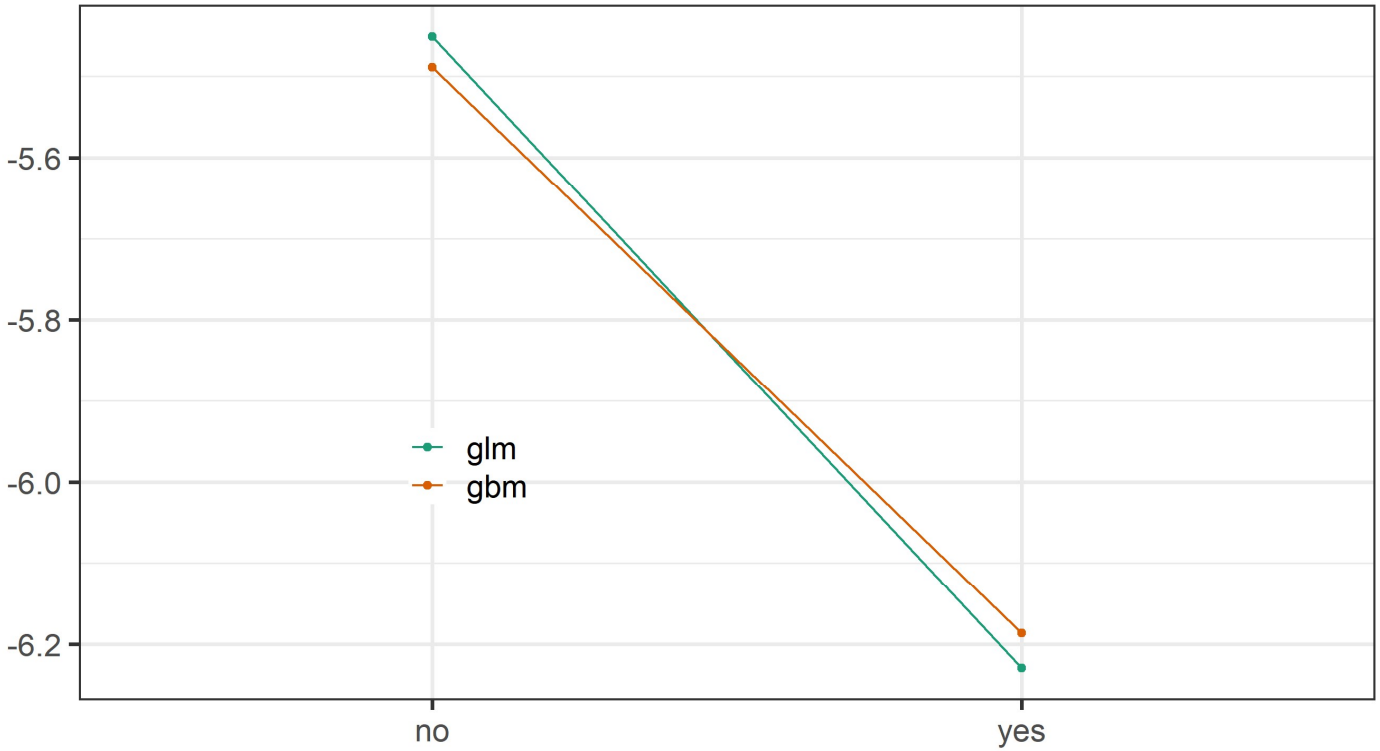
- For important variables, go through partial dependence plots.
- Here we select just some of them for illustration.
- On log-prediction scale in order to spot non-linearities.



Insights for driver's age

- Makes sense
- Invest more parameters

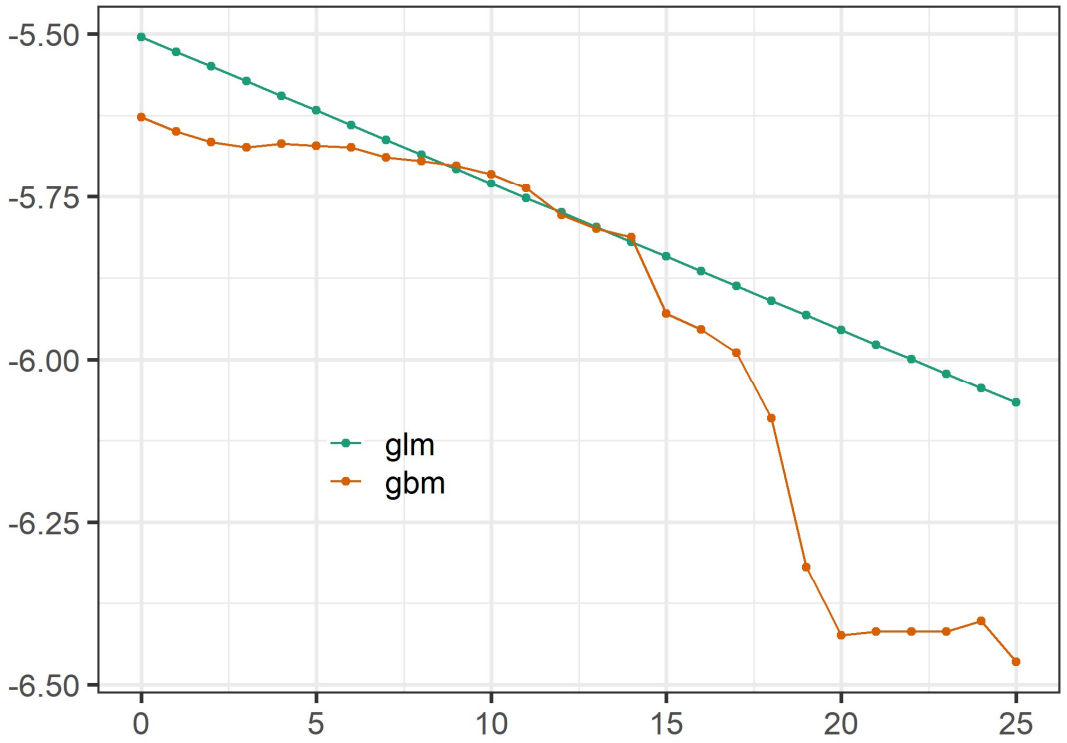
Main Effect for "Town (y/n)"?



Insights?

- Makes sense
- No action needed.

Main Effect for "Age of car"?



Insights?

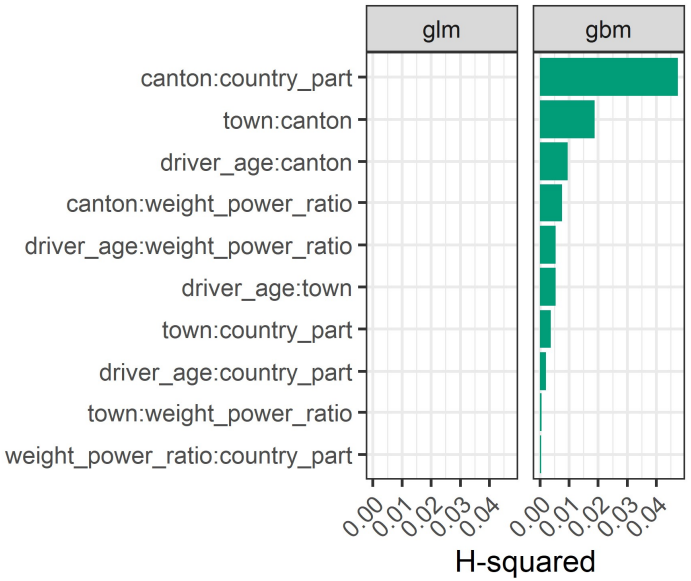
- Makes sense
- Add squared term.

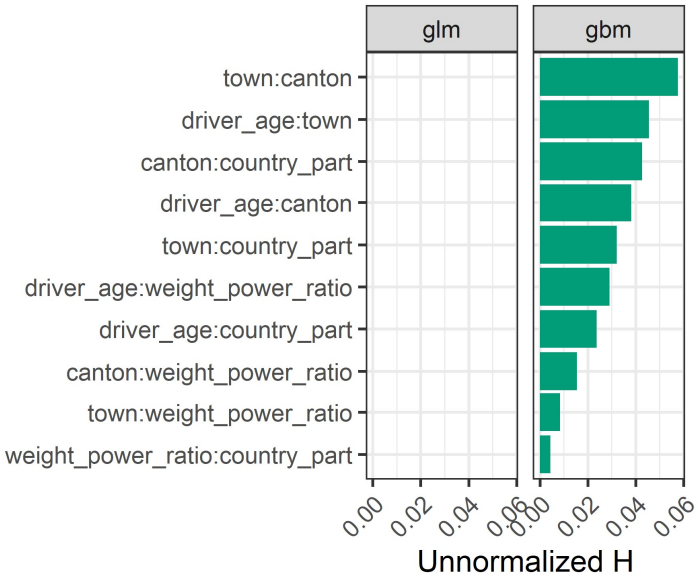
Interactions Required?

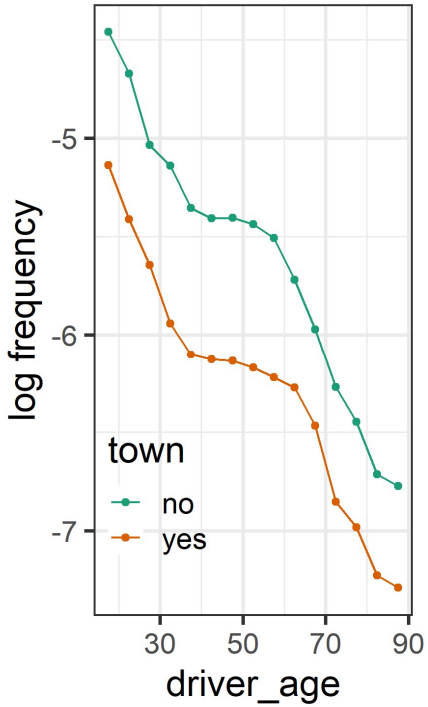
Friedman's H
based on centered partial dependence

$$H^2 = \frac{\sum\left(PD(x,z) - PD(x) - PD(z)\right)^2}{\sum PD(x,z)^2}$$

- Proportion unexplained by main effects.
- Evaluated on log-scale
- Weaknesses?





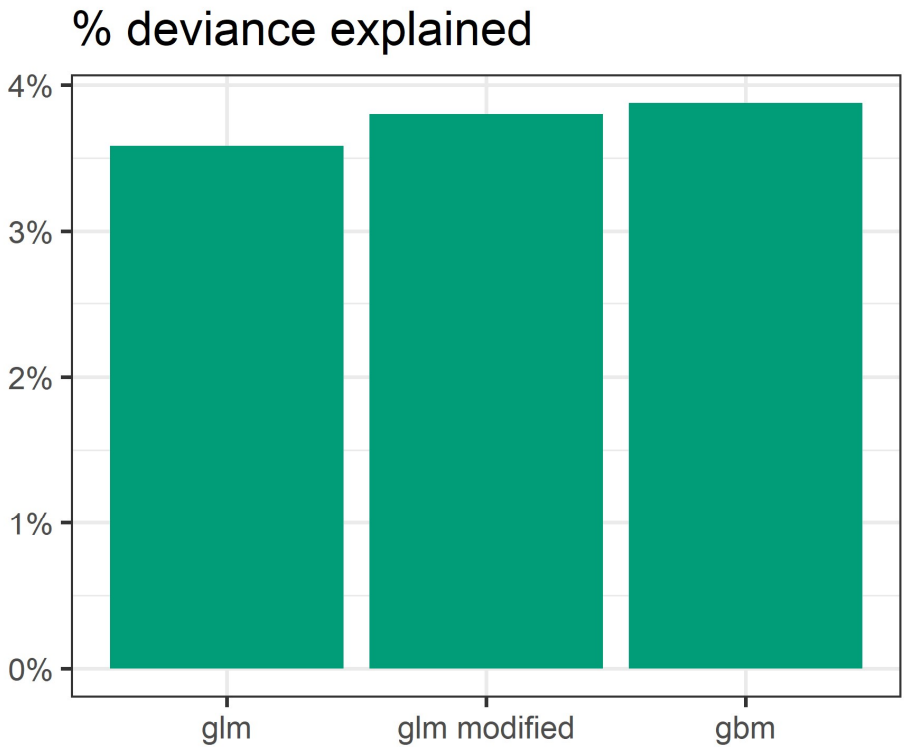


- Interpretation?
- No interactions!

Modified GLM Virtually as good as GBM!

Modifications

- Dropped two variables
- No interactions added
- Squared term for vehicle age
- Polynomial terms for driver age



Hurray! Trade-off works

The Future of XAI & Modern ML in Non-Life-Pricing

The Future of XAI & Modern ML in Non-Life-Pricing

In the long run, XAI allows to move from classic GLMs to modern ML.

In the short run, the proposed trade-off allows
to build GLMs in an efficient and guided way.

Resources

Resources

R-Packages used

- `tidyverse`: Data preparation & visualization
- `glmnetUtils`: Poisson-GLM
- `lightgbm`: Tree-Booster
- `MetricsWeighted`: Poisson-deviance
- `flashlight`: ML interpretations

R-Packages with publicly available pricing data sets

- `CASdatasets`: http://cas.uqam.ca
- `insuranceData` (CRAN)

Book on classic non-life insurance pricing

Ohlsson & Johansson (2010). Non-Life Insurance Pricing with Generalized Linear Models, Springer.