Non-Life Insurance Pricing goes XAI

Michael Mayer, October 6, 2021
Outline

1. Non-Life Insurance Pricing
2. Illustration with Real Data
3. The Role of ML and XAI
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?

<table>
<thead>
<tr>
<th>Response</th>
<th>Meaning</th>
<th>Distribution</th>
<th>Classic Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Claim cost</td>
<td>Losses ($$) / Exposure</td>
<td></td>
<td>???</td>
</tr>
</tbody>
</table>

\[
\text{Claim Cost} = \frac{\text{Losses}}{\text{Exposure}} = \frac{\text{Claim Count}}{\text{Exposure}} \cdot \frac{\text{Losses}}{\text{Claim Count}}
\]

<table>
<thead>
<tr>
<th>Response</th>
<th>Meaning</th>
<th>Distribution</th>
<th>Classic Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>Claim count / Exposure</td>
<td></td>
<td>Poisson GLM with log-link</td>
</tr>
<tr>
<td>Severity</td>
<td>Losses / Claim count</td>
<td></td>
<td>Gamma GLM with log-link</td>
</tr>
</tbody>
</table>

Challenges?
2. Illustration with Real Data
Car Collisions with Large Animals

Annual figures for Switzerland

- Many large animals die on Swiss roads, among them ~8'000 deers.
- 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
- Claim frequency around 0.3%-0.5%

Model for claim frequency, taking into account individual risk factors?
A simple GLM for Claim Frequency

<table>
<thead>
<tr>
<th>Data</th>
<th>Model*</th>
</tr>
</thead>
</table>
| • Car policies with partial/full coverage  
• Data over multiple years  
• Millions of data rows  
• Train/test split grouped by policy | • Poisson-GLM with log-link  
\[ \log E(y) = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m \]  
• Estimates of \( \beta \) minimize deviance  
\[ 2 \sum w_i (y_i \log(y_i / \hat{y}_i) - (y_i - \hat{y}_i)) \]  
• Some estimates*:  
  • Driver_age: \(-0.03\)  
  • Town: \(-0.80\)  
  • max_7000_km: \(-0.66\) |

*Toy model for illustration only

Features

- **Driver**: place of living, age, gender, …
- **Car**: price, age, weight-to-power ratio, leased, …
- **Policy**: bonus protection, fully or partially comprehensive, year, …
3. The Role of ML and XAI
General View: Insurance Pricing Today and in Future

Today

Classic modelling
Manually build GLM

Difficult to select features, interactions, non-linearities → slow model building

Trade-off
Use insights from modern ML & XAI to build GLM

Efficient model building

Future?

Modern ML & XAI?

- Automatic model building
- Scales (more models, …)

Illustrated for toy model with the help of Gradient Boosting and SHAP

Advantages of GLMs?
1. Ease of interpretation and info gain
2. Transparency
3. Industry standard
4. Well-supported by standard software

Non-Life Insurance Pricing goes XAI 06.10.2021
What is Gradient Boosting?

A Simple Decision Tree

Is it raining?
- Yes: Umbrella
- No: Bad weather forecast?
  - Yes: Umbrella
  - No: No umbrella

Questions can be derived from data

Gradient Boosting
1. Fit small decision tree
2. Correct errors by another small tree
3. Repeat Step 2 multiple times

Implementations
- XGBoost
- LightGBM
- CatBoost

Click for more comparisons...
What is SHAP?

**Fair, additive decomposition of single prediction**
- Game-theoretic approach based on Shapley values
- Lundberg & Lee (2017)

**Variable Importance**
- Note: this is just one way to measure variable importance
- `空 | SHAP value` per feature

**Dependence Plots**
- Vertical scatter indicates interaction effects
- `空 | SHAP value` of feature against feature value

---

**SHAP values**
- 1 value per prediction and feature

**Value Breakdown**
- `0.7`
Performance on 20% Test Data?

Relative reduction in Poisson deviance loss

- Makes sense: animal collisions are quite random and cannot be well predicted.
- There is room for improvement for GLM, but not too much.

Insights
Important Features?

- Results make sense → trust in models grow
- Same variables important across models

Insights
Feature Effects?

Example: Driver's age

Example: Town (yes / no)

Insights
- Effects make sense
- Similar across models
- Use additional parameters for "driver's age" in GLM
- Add interaction of "town" and "max_7000_km" to GLM

Action summary
- Add parameters to GLM in guided way by XAI to reduce performance gap
- Go for GLM
Key Takeaways
Key Takeaways

- GLMs stay important in non-life insurance pricing.
- ML + XAI is a great way to improve them.
Resources


Code examples:
- [github.com/slundberg/shap](https://github.com/slundberg/shap)
- [https://github.com/mayer79/python_notebooks](https://github.com/mayer79/python_notebooks)
- [https://github.com/JSchelldorfer/ActuarialDataScience (Nb 8)](https://github.com/JSchelldorfer/ActuarialDataScience (Nb 8))