Modeling lapse rates using machine learning

Ulm Actuarial Session - Convention A

- Lucas Reck
- September 19th, 2022
- joint work with Andreas Reuß and Johannes Schupp
**Introduction**

Motivation for a lapse model

- Lapse risk is one of the key risk drivers of life business.
  - significant impact on the cash flow profile and the profitability of life insurance business
  - relevant for Asset-Liability-Management and liquidity risk
- Market consistent valuations are based on best estimate future lapse rates.
  - e.g. Solvency II regulation (also specific risk module that addresses lapse risk)
Introduction
Common practice

- Whittaker-Henderson (univariate smoothing algorithm)
- Prespecified covariate (e.g. contract duration)
Introduction
Problem of the common practice

- Whittaker-Henderson including covariate country

- The insurance portfolio is typically divided into sub-portfolios based on contract characteristics like type of contract, country, or distribution channel.
Introduction
Motivation for the Lasso

- Multivariate models - using all covariates simultaneously.
- GLM lapse model: Eling and Kiesenbauer (2014) and Barucci et al. (2020)
  - number of coefficients → considerable effort
  - risk of under- or overfitting
- Data Science methods can be a solution. We use the Lasso approach to derive a lapse model that
  - is calibrated automatically and purely data driven,
  - but remains fully interpretable,
  - is able to detect hidden structures in the covariates.
- We analyze and combine different extensions of Lasso to satisfy the needs of a practical application.
Introduction

Data set

Application

We use data from a European life insurer operating in four countries (run-off portfolio).

We use 13 covariates and a total sample size of 501,251.

Covariates include standard data of an insurance company, e.g.:
- contract duration, entry age, sum insured, country, contract type,...
**Method**

Logistic regression

- Logistic regression
  - $Y_i$ is Bernoulli distributed.
  - $E(Y_i) = p(x_i)$
  - Transform $p(x_i)$ and assume a linear relationship:
    \[
    \text{logit}(p(x_i)) = \ln \left( \frac{p(x_i)}{1 - p(x_i)} \right) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_m x_{im}
    \]
  - Likelihood function:
    \[
    L(\beta, X, y) = \prod_{i=1}^{n} p(x_i)^{y_i} (1 - p(x_i))^{(1 - y_i)}
    \]
Method
Lasso

Lasso (Least Absolute Shrinkage and Selection Operator)

- Include a regularisation term:

\[
\min - \log(L(\beta, X, y)) + \lambda \sum_{j=1}^{J} g(\beta_j)
\]

**Shrinkage-Factor: \( \lambda \geq 0 \)**
Controlling the impact of regularisation and goodness-of-fit

**Regularisation:**
Penalty term for the coefficients
Regular Lasso: \( g(\beta_j) = \sum_{i=1}^{p_j} |\beta_{j,i}| \)
Method
Extension: Fused Lasso and Trend Filtering
Tibshirani and Taylor (2011)

- Now we extend the Lasso: \( \min \, - \log(L(\beta, X, y)) + \lambda \sum_{j=1}^{J} g_j(\beta_j) \)

- Regular Lasso: \( g_R(\beta_j) = \| \beta_j \|_1 = \sum_{i=1}^{p_j} |\beta_{j,i}| \)

- Fused Lasso:
  \[
  g_F(\beta_j) = \sum_{i=2}^{p_j} |\beta_{j,i} - \beta_{j,i-1}| =: \sum_{i=2}^{p_j} |\beta_{j,i}^F|
  \]

- Trend Filtering:
  \[
  g_T(\beta_j) = \sum_{i=3}^{p_j} |\beta_{j,i} - 2\beta_{j,i-1} + \beta_{j,i-2}| =: \sum_{i=3}^{p_j} |\beta_{j,i}^T|
  \]
Model Selection
Preparation

- R interface for H2O
- Assign a penalty term for each covariate:
  - Contract duration → trend
  - Entry age → fused
  - Sum insured → trend
  - Country → regular
  - ...
- Hyperparameter $\lambda$ is based on 5-fold cross validation with one standard error rule.
- Residual Deviance as measure for goodness of fit
Model Selection
Trend filtering for contract duration
Model Selection
Fused Lasso for entry age

![Graph showing lapse rates and number of observations across different entry age groups.](image)

- Lapse rate
- Entry age
- Number of observations

\[ \lambda_{\text{max}} \]

\[ \lambda \]

© September 2022  Modeling lapse rates using machine learning
Interactions

Motivation – Problem of the model without interactions

- Impact of contract duration differs for the individual countries
- Model without interaction does not capture this
- We want to include the interaction contract duration - country
Interactions
Model with the interaction contract duration - country

![Heatmap of contract duration and country interactions](image)

- Observed
- Prediction
- Marginal

$\lambda_{max}$ $\lambda$ contract duration

© September 2022  Modeling lapse rates using machine learning
Conclusion
Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of parameters</th>
<th>1 - Deviance/Null Deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept Only</td>
<td>1</td>
<td>0%</td>
</tr>
<tr>
<td>Whittaker-Henderson</td>
<td>20</td>
<td>6.7%</td>
</tr>
<tr>
<td>Lasso without interaction</td>
<td>44 (out of 77)</td>
<td>12.1%</td>
</tr>
<tr>
<td>Lasso with interaction</td>
<td>79 (out of 145)</td>
<td>12.9%</td>
</tr>
</tbody>
</table>

Advantages - The resulting model

- is multivariate and estimates lapse rates using all covariates simultaneously,
- is calibrated automatically and purely data driven,
- remains fully interpretable,
- is able to detect hidden structures in the covariates.
Conclusion
Further results and outlook for future research

- Sensitivity analysis:
  - Base Model
  - “Screening” vs “Selecting” property of the Lasso
  - Penalty types
  - Macroeconomic covariates
  - Elastic net approach
  - Offset model for interactions

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of parameters</th>
<th>$1 - \frac{D}{D_0}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lasso without interaction</td>
<td>44</td>
<td>12.1%</td>
</tr>
<tr>
<td>“Screening” Lasso</td>
<td>30</td>
<td>12.1%</td>
</tr>
<tr>
<td>Lasso all regular</td>
<td>70</td>
<td>12.2%</td>
</tr>
<tr>
<td>Macroeconomic</td>
<td>72</td>
<td>13.3%</td>
</tr>
<tr>
<td>Elastic net, $\alpha = 50%$</td>
<td>55</td>
<td>12.2%</td>
</tr>
<tr>
<td>Offset model</td>
<td>64</td>
<td>12.7%</td>
</tr>
</tbody>
</table>

- Outlook for future research
  - Other machine learning approaches (random forest, neural networks, etc.)
  - Multistate model (active, paid-up, lapse)


Contact

Johannes Schupp
+49 731 20644-241
j.schupp@ifa-ulm.de

Lucas Reck
+49 731 20644-239
l.reck@ifa-ulm.de

Andreas Reuß
+49 731 20644-251
a.reuss@ifa-ulm.de