Multistate analysis of policyholder behaviour in life insurance - Lasso based modelling approaches

IME Congress 2023

- Lucas Reck

- July 5th, 2023

- joint work with Andreas Reuß and Johannes Schupp

Agenda

Introduction

Modelling approaches

Real world application

Comparison and results

Conclusion

References

Introduction
Motivation

- Different states and transitions for life insurance policies:
  - Active, paid-up, reinstatement, lapse, death, etc.
  - Affect the cash flow profile and therefore the ALM → Solvency II
  - In practice, independent (binary) models are built to describe a certain effect, but typically no holistic model set-up
- Different modelling approaches are used model multi-class situations:
  - Survival analysis
  - Machine learning approaches (Random forest, GBM, etc.)
  - Generalised Linear Models (GLM)
- We choose different GLM based approaches with the Lasso penalisation to derive a model which
  - is calibrated automatically and purely data driven,
  - but remains fully interpretable,
  - is able to detect hidden structures in the covariates.

Modelling approaches

Multi-class situation

- Two ways of dealing with a multi-class situation:
  - Decomposition strategies
    - One vs. all (OVA)
    - One vs. one (OVO)
    - Nested models
  - Holistic approach
    - Multinomial logistic regression (MLR)
- Different ways of including the transition history
  - No inclusion
  - Markov property (using the previous state)
    - As a covariate
    - As a covariate including its interaction terms
    - By splitting the data set
  - Full transition history (using the time since being paid-up)

Modelling approaches

One vs. all (OVA)

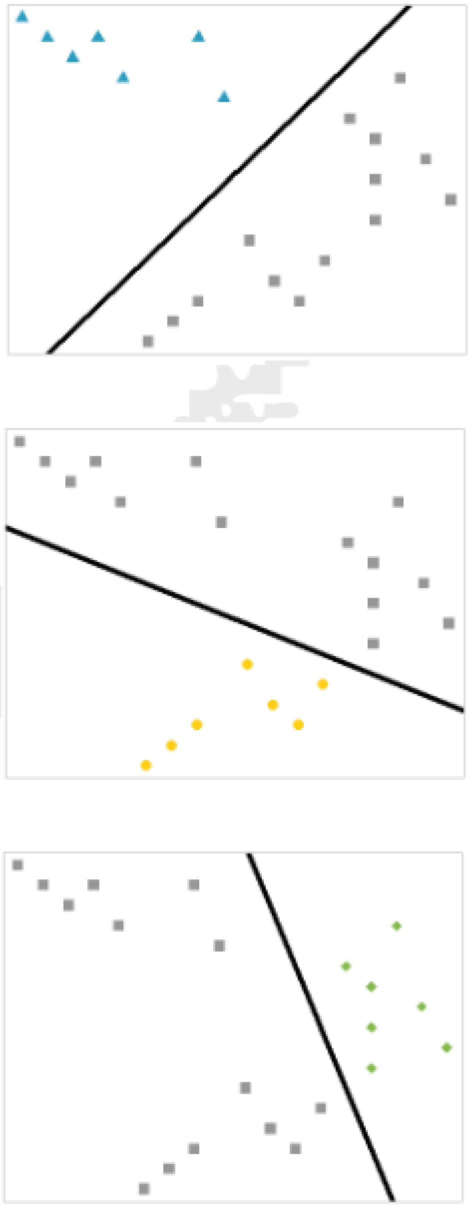
- Models one class versus all other classes:

$$M = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix}$$

- Aggregation:

$$q_k = \frac{p_k}{\sum_i p_i}$$

- In general, there are m independent models



Modelling approaches

One vs. one (OVO)

- Models one class versus another class:

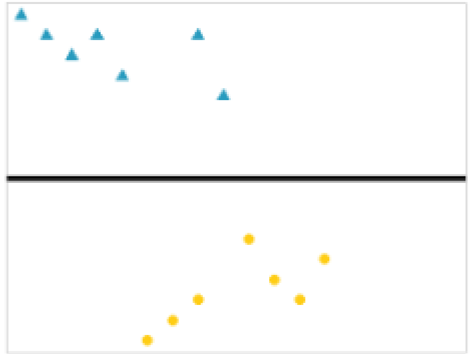
$$M = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & -1 \end{bmatrix}$$

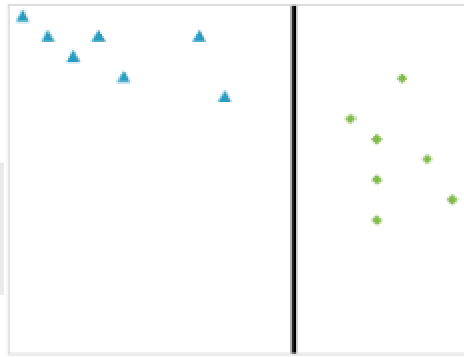
- Aggregation:

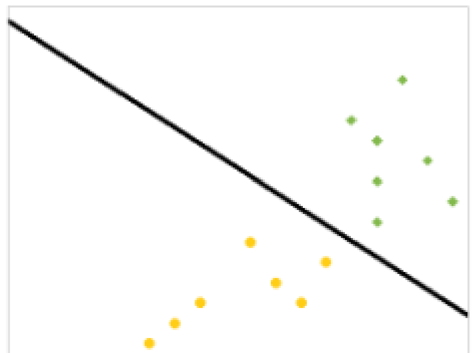
  Minimised (weighted) sum of Kullback-Leibler distances between

  $P(Y = k | Y = k \text{ } or \text{ } Y = j)$ and $q_k = \frac{q_k}{q_k + q_j}$

- In general, there are $\frac{m(m-1)}{2}$ independent models





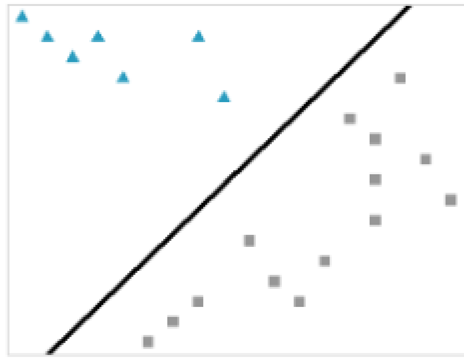


Modelling approaches

Nested approach

- Models in a hierarchical order

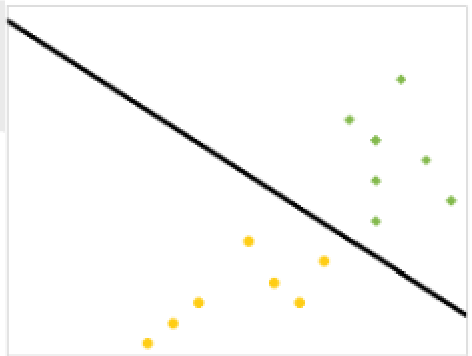
$$\mathrm{M} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ -1 & -1 \end{bmatrix} \text{ or alternatively } \mathrm{M} = \begin{bmatrix} -1 & 1 \\ 1 & 0 \\ -1 & -1 \end{bmatrix} \text{ or } \mathrm{M} = \begin{bmatrix} -1 & 1 \\ -1 & -1 \\ 1 & 0 \end{bmatrix}$$



- Aggregation:

  According to the corresponding path, e.g.:

$$P(Y = C) = P_1(Y = B \text{ } or \text{ } Y = C) * P_2(Y = C | Y = B \text{ } or \text{ } Y = C)$$



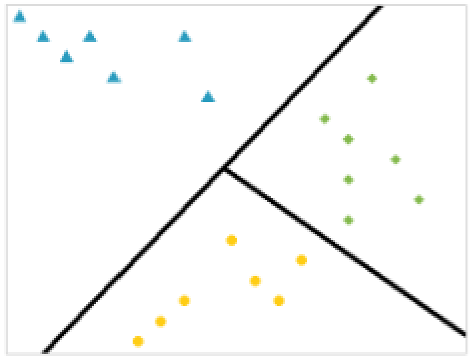
- In general, there are $m - 1$ independent models,

  but $\mathcal{O}(2^m m!)$ different orders

Modelling approaches

MLR

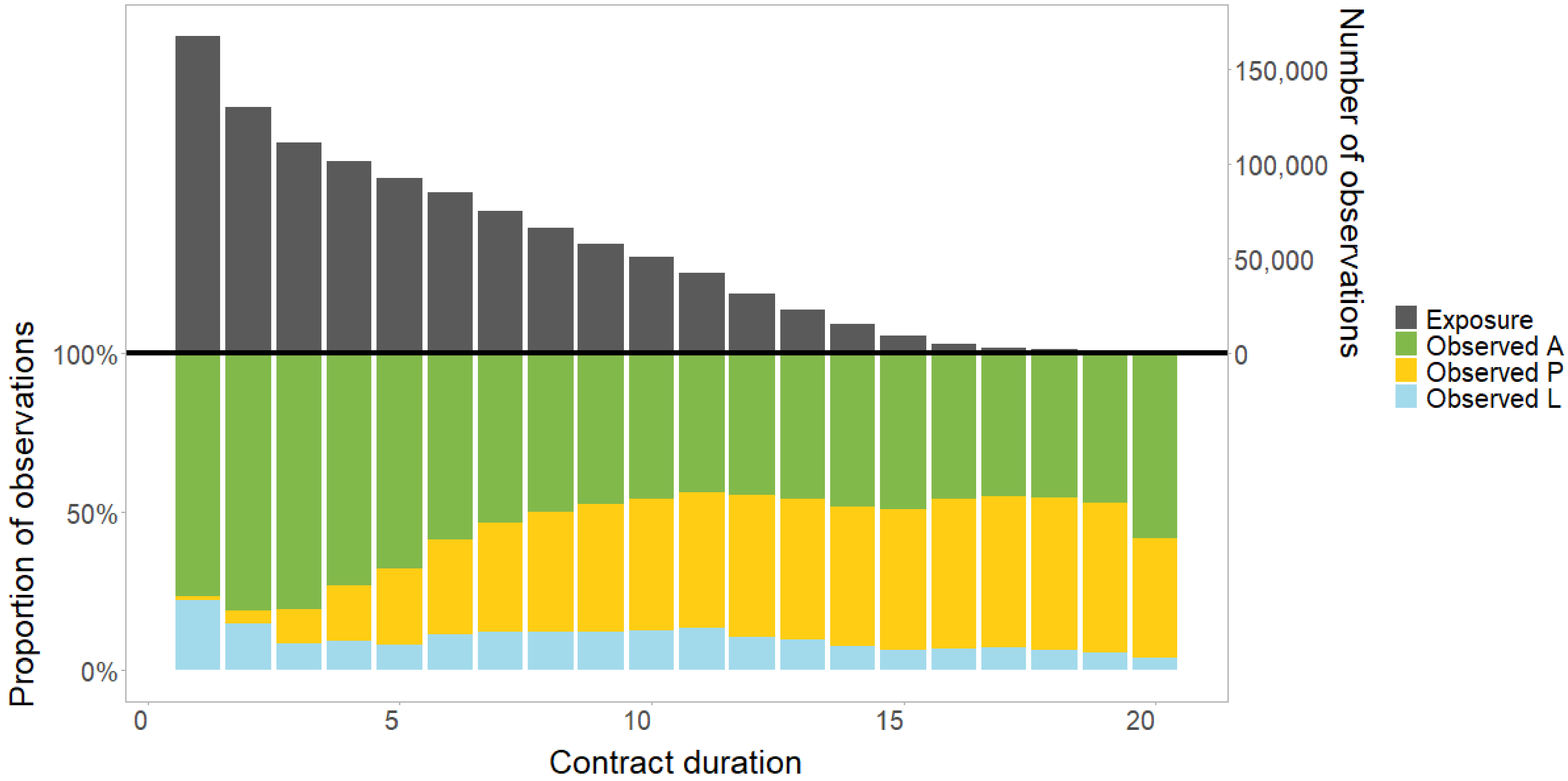
- No decomposition into several independent binary models
- No aggregation
- Exactly 1 model



Real world application

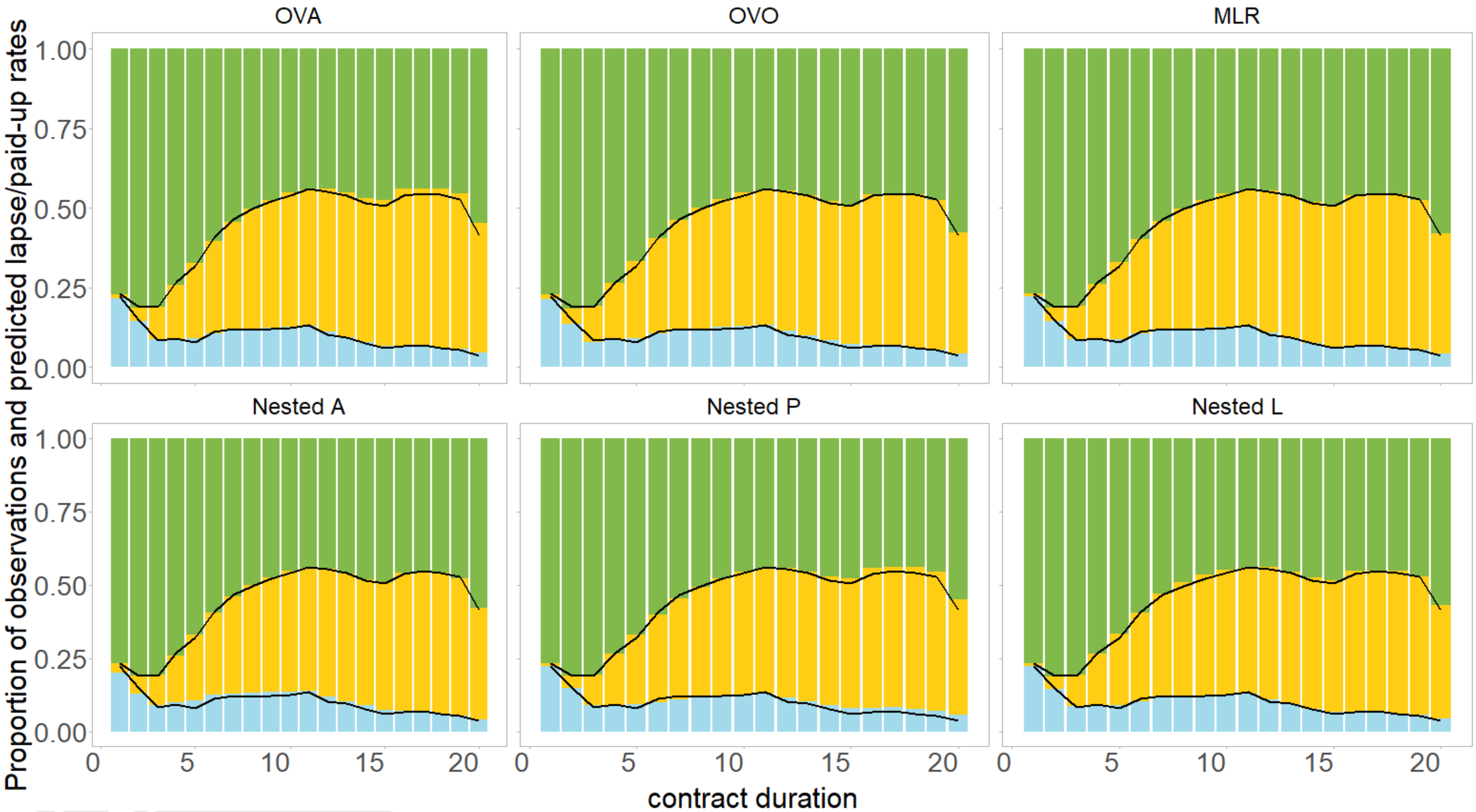
- Data set:
  - 21 years observation period
  - Around 1 million observations from 170k unique contracts
  - 15 covariates
  - 3 states: active, paid-up, lapse (no reinstatement)
- Implementation uses a R interface for H2O
- Assign a (extended) Lasso penalty term for each covariate:
  - Contract duration → trend
  - Entry age → fused
  - Sum insured → trend
  - Country → regular
  - …
- Hyperparameter λ is based on 5-fold cross validation with one standard error rule.
- Residual Deviance as measure for goodness of fit

Real world application



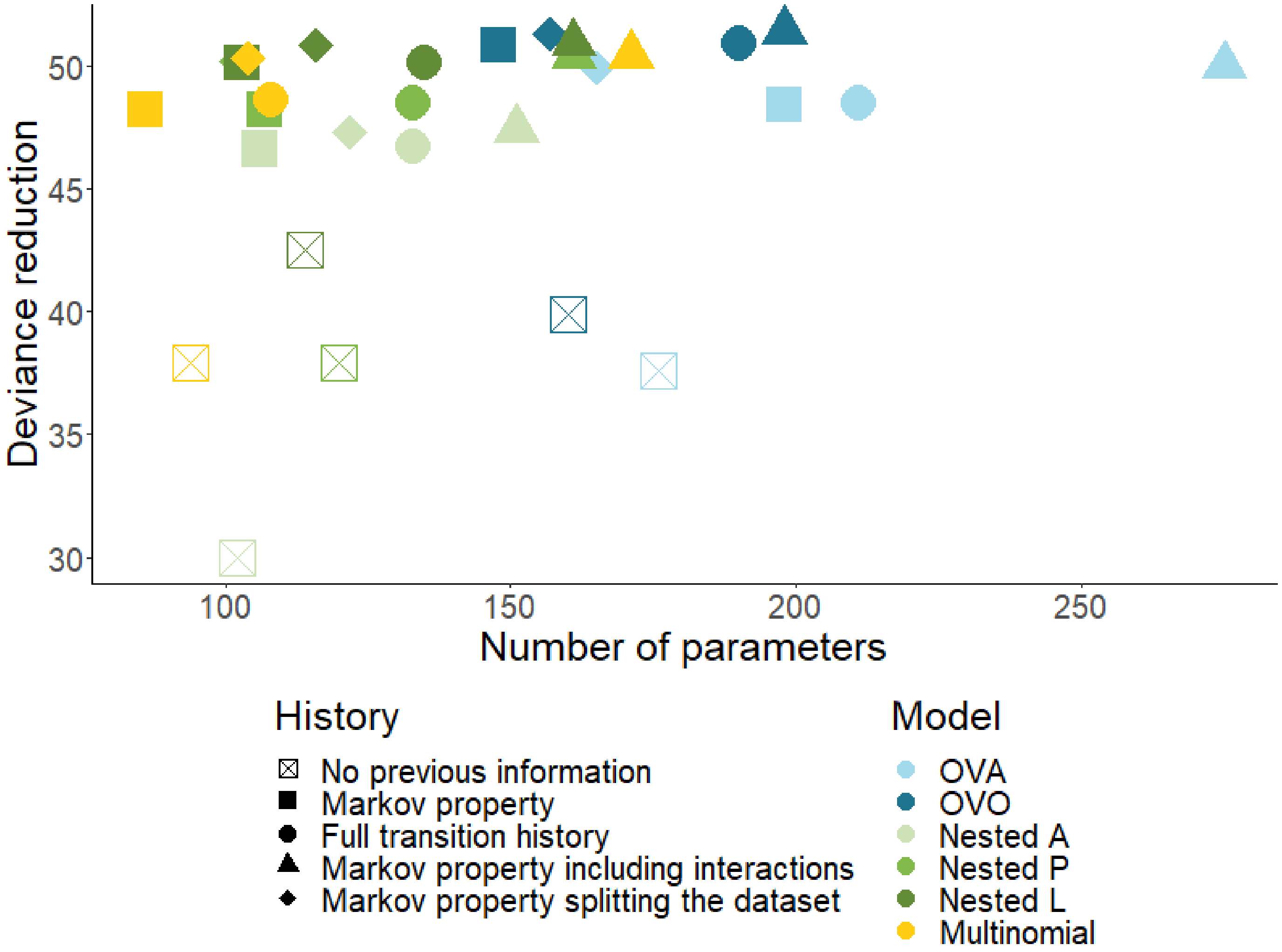
Comparison and results

Overall prediction for different values of contract duration



Comparison and results

Number of parameters and deviance reduction



Conclusion
Transition history

- Previous state has significant impact on model performance
- Full transition history does not seem to add value to the models
- Including the previous state with its interactions improves the model slightly, but the number of parameters increases accordingly
- Splitting the data set performs slightly better than including the previous state as a covariate. However, the number of models increases, and it might be unfeasible for more states

Conclusion
Modelling approach (quantitatively)

- Overall, model performances are on a similar level, but generally:
  - 1) OVO
  - 2) Nested L
  - 3) MLR, OVA and Nested P
  - 4) Nested A
- In terms of the number of parameters:
  - 1) MLR
  - 2) Nested A
  - 3) Nested P, Nested L
  - 4) OVO
  - 5) OVA

Conclusion

Modelling approach (qualitatively)

- OVO is hard to interpret due to its complicated aggregation scheme
- Nested approach has a lot of different definitions (especially for a large number of classes)
- Overall, the MLR has the most qualitative advantages:
  - unique definition with one model
  - easy to interpret
  - easy to generalise

References

- Barucci, E., Colozza, T., Marazzina, D. and Rroji, E. (2020). The determinants of lapse rates in the Italian life insurance market. European Actuarial Journal, 10(1), 149-178. https://doi.org/10.1007/s13385-020-00227-0
- Eling, M. and Kochanski, M. (2013). Research on lapse in life insurance: what has been done and what needs to be done?. The Journal of Risk Finance, 14(4), 392-413. https://doi.org/10.1108/JRF-12-2012-0088
- Hastie, T. and Tibshirani, R. (1997). Classification by pairwise coupling. Advances in neural information processing systems, 10. https://proceedings.neurips.cc/paper/1997/file/70feb62b69f16e0238f741fab228fec2-Paper.pdf
- Lorena, A.C., De Carvalho, A.C.P.L.F. and Gama, J.M.P. (2008). A review on the combination of binary classifiers in multi-class problems. Artificial Intelligence Review, 30(1), 19-37. https://doi.org/10.1007/s10462-009-9114-9
- Milhaud, X. and Dutang, C. (2018). Lapse tables for lapse risk management in insurance: a competing risk approach. European Actuarial Journal, 8(1), 97-126. https://doi.org/10.1007/s13385-018-0165-7
- Reck, L., Schupp, J. and Reuß, A. (2022). Identifying the determinants of lapse rates in life insurance: an automated Lasso approach. European Actuarial Journal, 1-29. https://doi.org/10.1007/s13385-022-00325-1

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