Institut für Finanz- und Aktuarwissenschaften

Modeling lapse rates using machine learning

Ulm Actuarial Session - Convention A

- Lucas Reck

- September 19th, 2022

- joint work with Andreas Reuß and Johannes Schupp

Agenda

Introduction

Method

Model Selection

Interactions

Conclusion

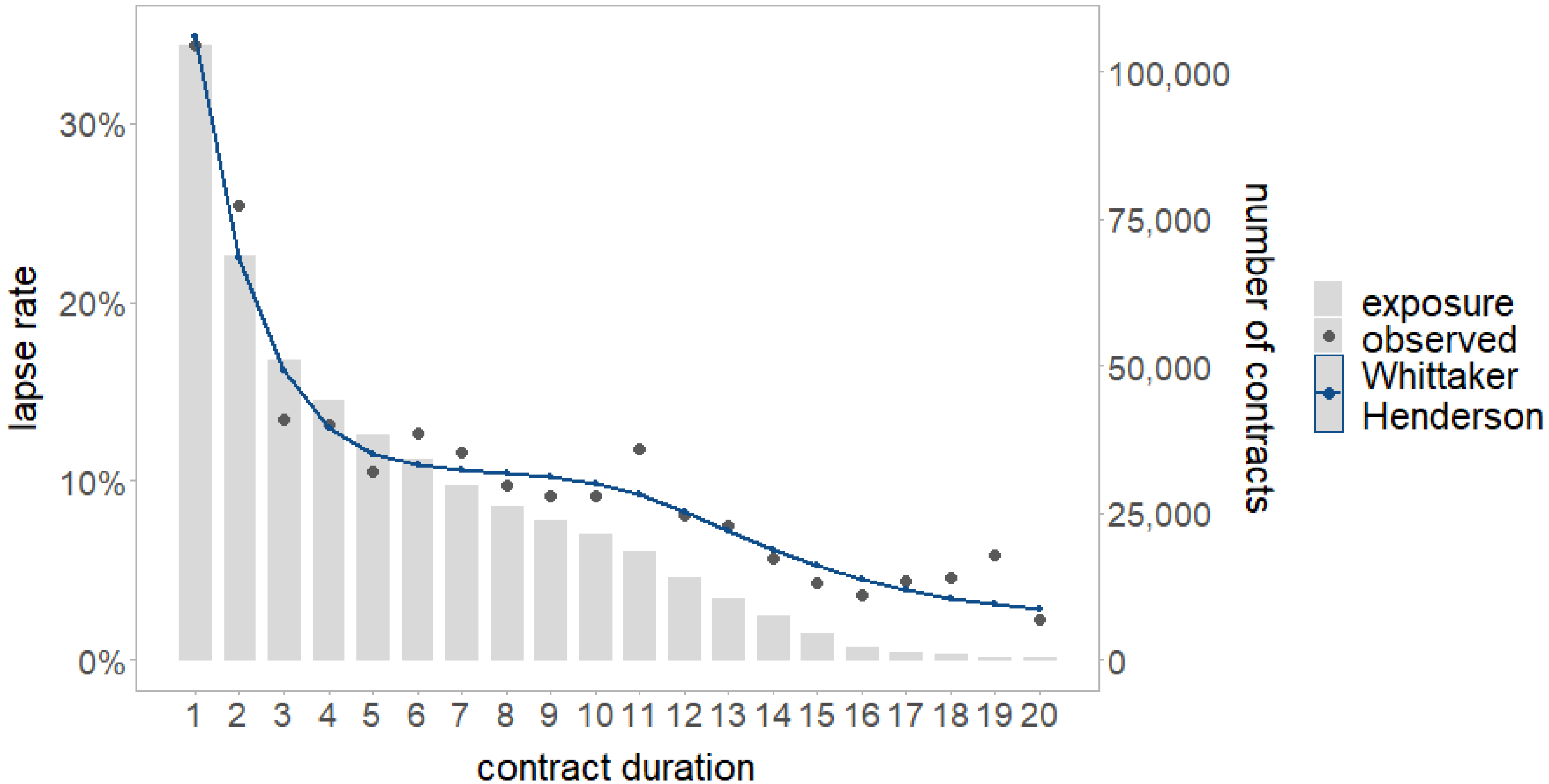
References

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

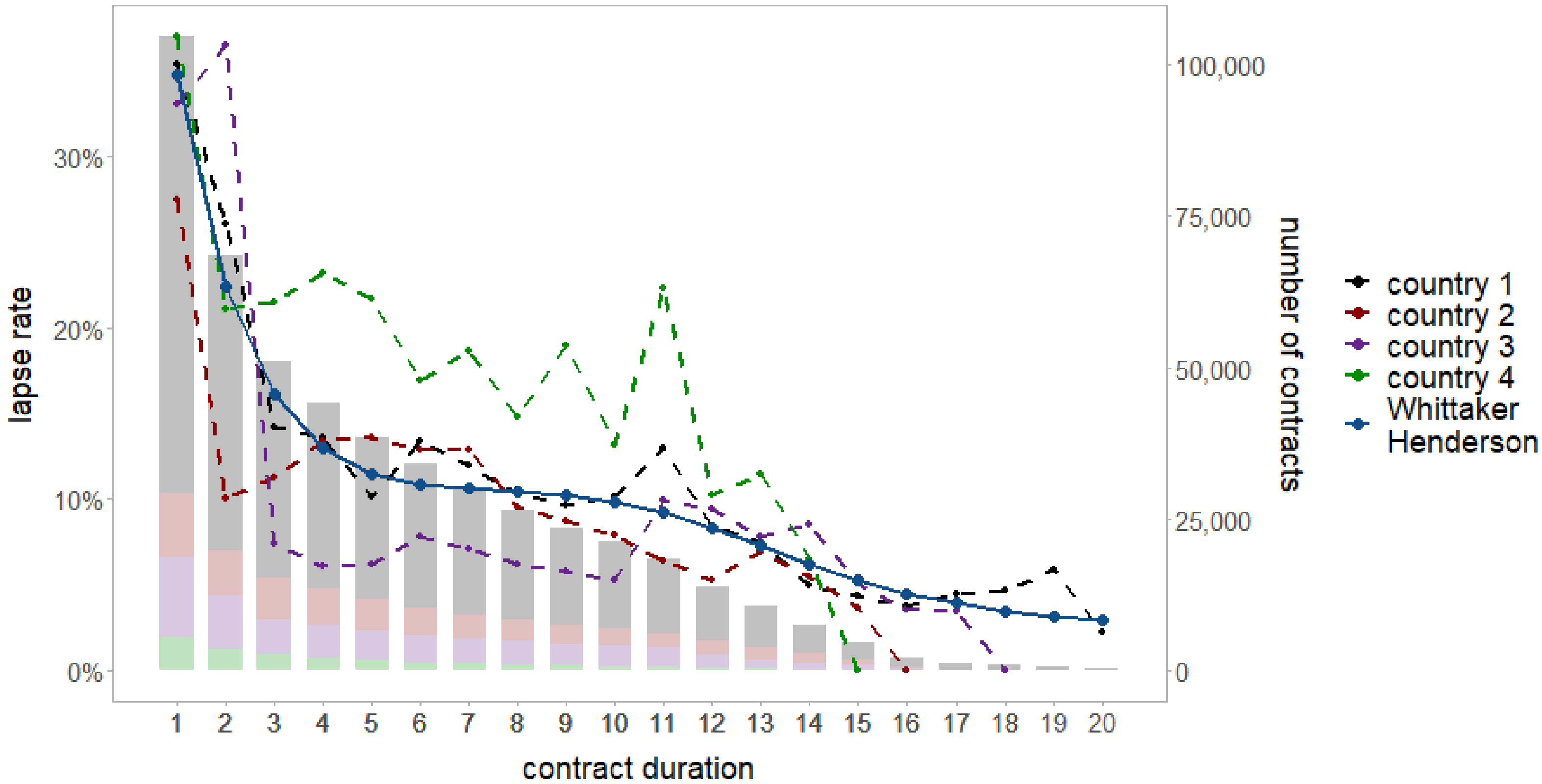
- Whittaker-Henderson (univariate smoothing algorithm)
- Prespecified covariate (e.g. contract duration)



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:

$$\text{logit}\big(p(x_i)\big) = \ln\left(\frac{p(x_i)}{1-p(x_i)}\right) = \beta_0 + \beta_1 x_{i1} + \ \dots + \beta_m x_{im}$$

  - Likelihood function:

$$L(\beta, X, y) = \prod_{i=1}^{n} p(x_i)^{y_i}(1 - p(x_i))^{(1-y_i)}$$

ifa

Method
Lasso

- Lasso (Least Absolute Shrinkage and Selection Operator)
  - Include a regularisation term:

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

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

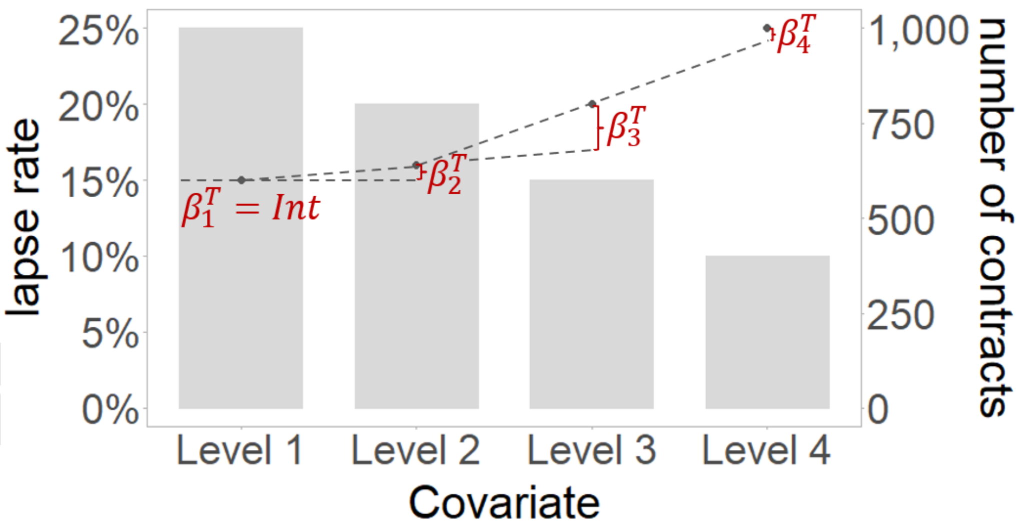
Extension: Fused Lasso and Trend Filtering

Tibshirani and Taylor (2011)

- Now we extend the Lasso: $\min \; -\log\big(L(\beta, X, y)\big) + \lambda \sum_{j=1}^{J} g_j(\beta_j)$
  - Regular Lasso: $g_R(\beta_j) = \| \beta_j \|_1 = \sum_{i=1}^{p_j} |\beta_{j,i}|$
  - Fused Lasso:
  - Trend Filtering:

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

$$g_T(\beta_j) = \sum_{i=3}^{p_j} |\beta_{j,i} - 2\beta_{j,i-1} + \beta_{j,i-2}| =: \sum_{i=3}^{p_j} |\beta_{j,i}^{\mathrm{T}}|$$

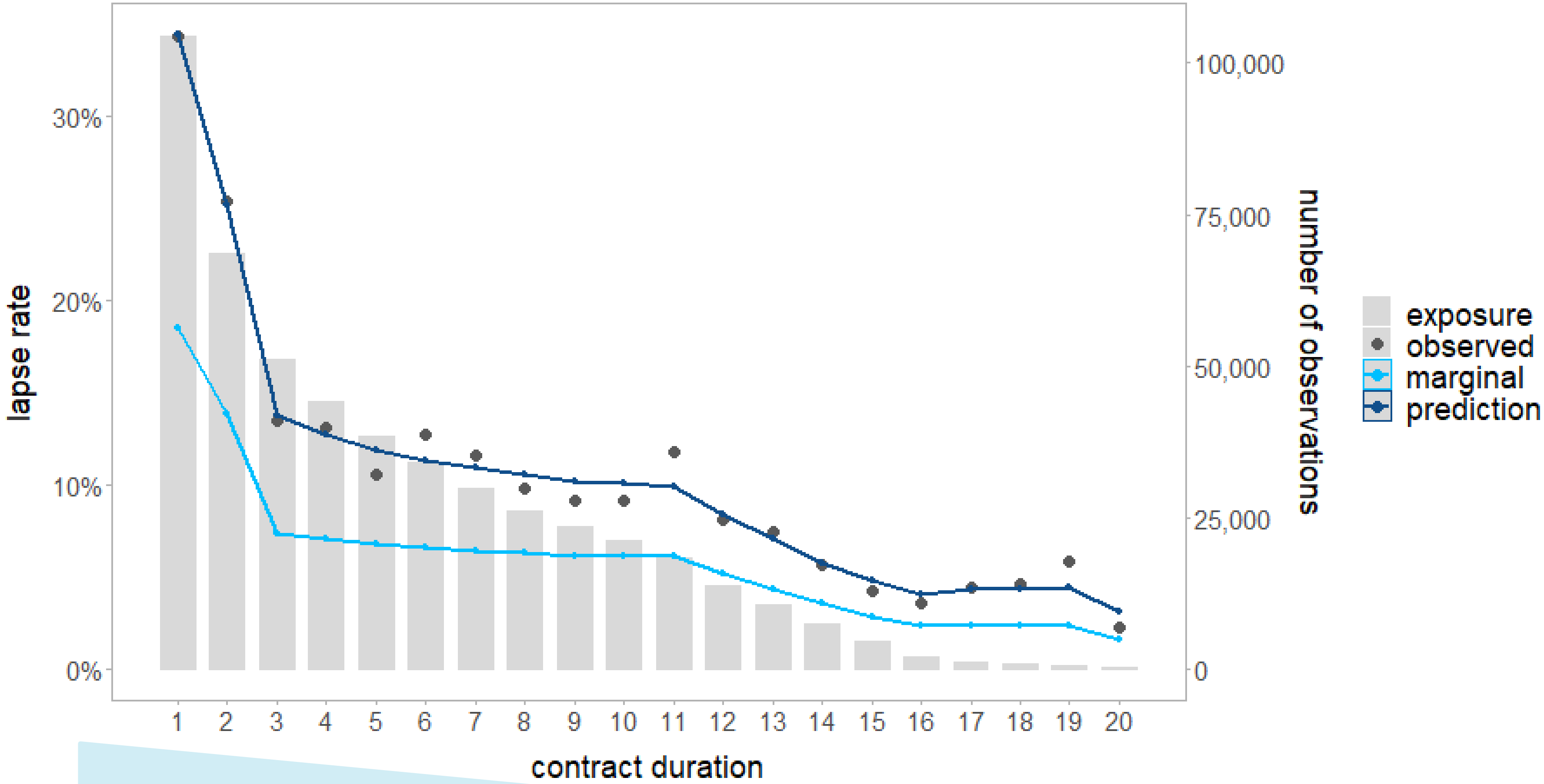


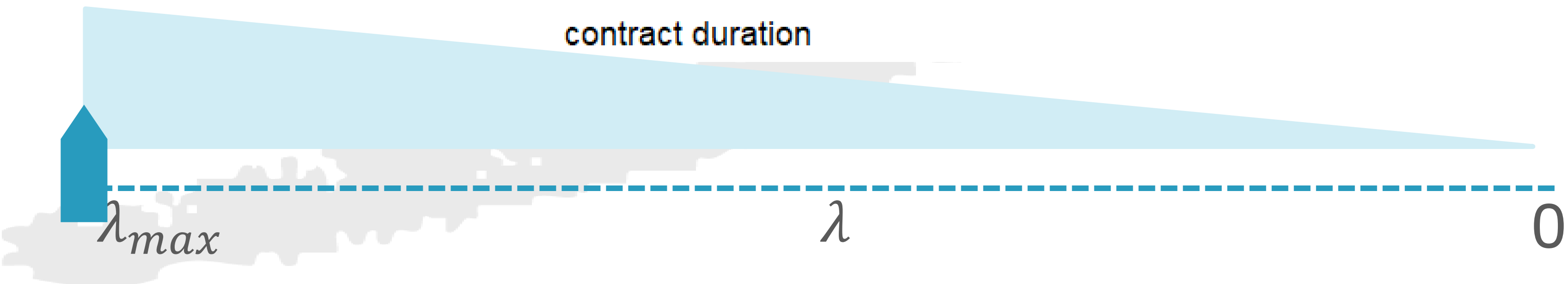
Model Selection

Preparation

- R interface for H2O
- Assign a penalty term for each covariate:
  - Contract duration → trend
  - Entry age → fused
  - Sum insured → trend
  - Country → regular
  - …
- Hyperparameter λ is based on 5-fold cross validation with one standard error rule.
- Residual Deviance as measure for goodness of fit

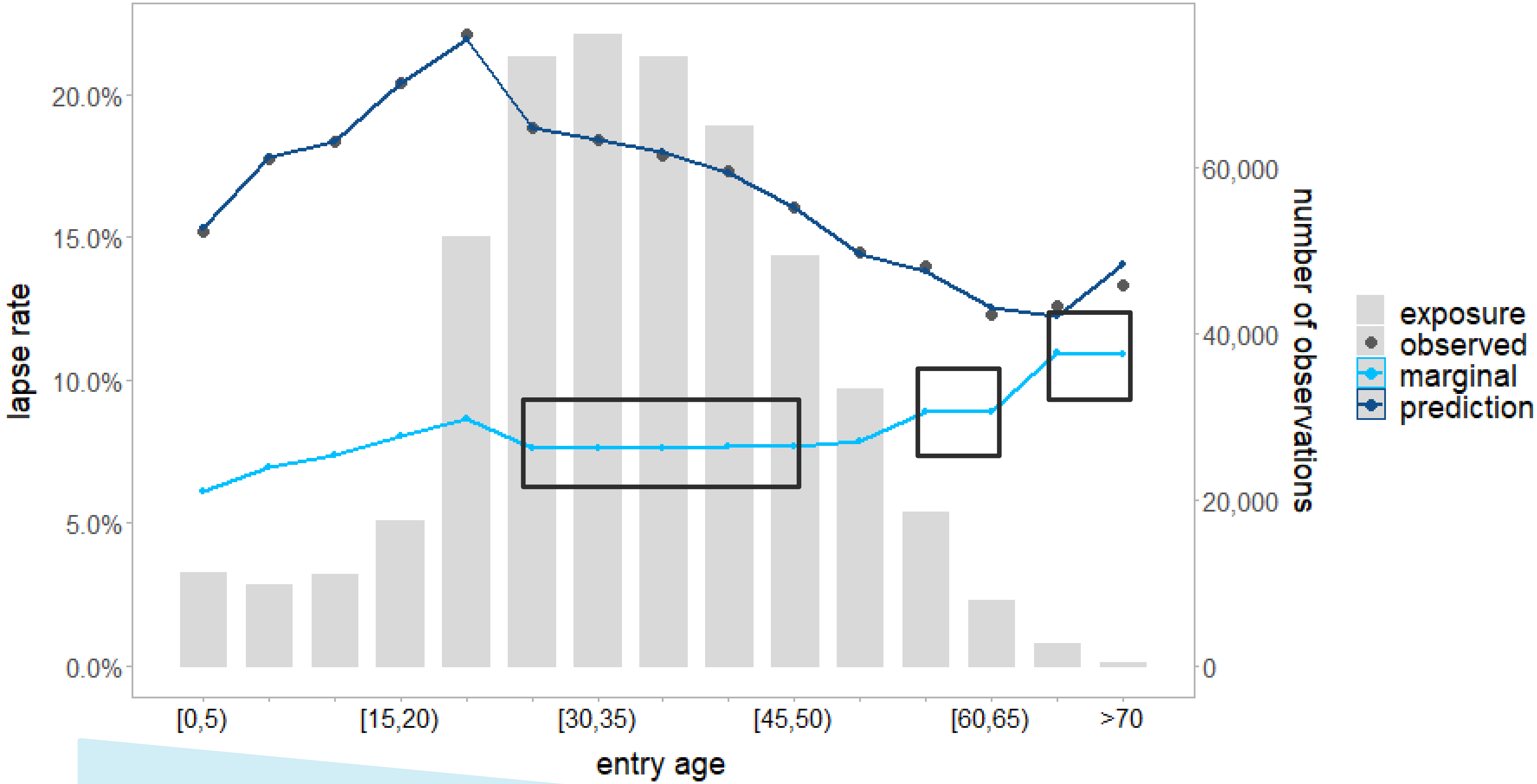
Model Selection
Trend filtering for contract duration

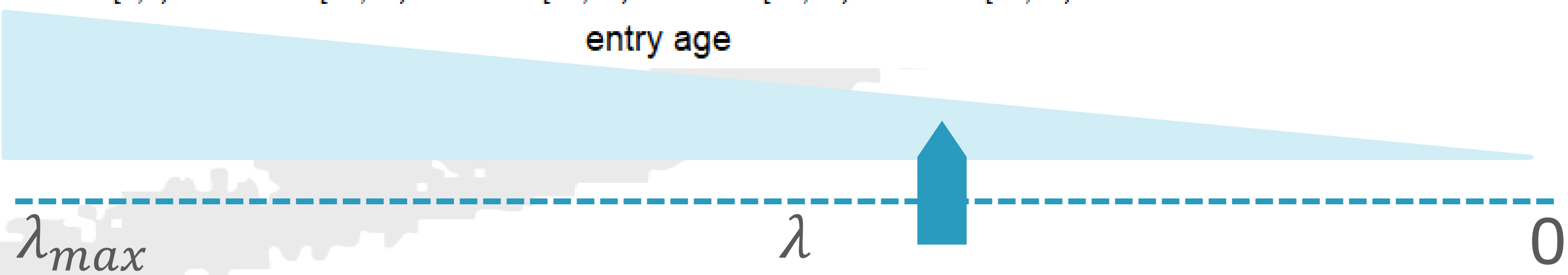




Model Selection

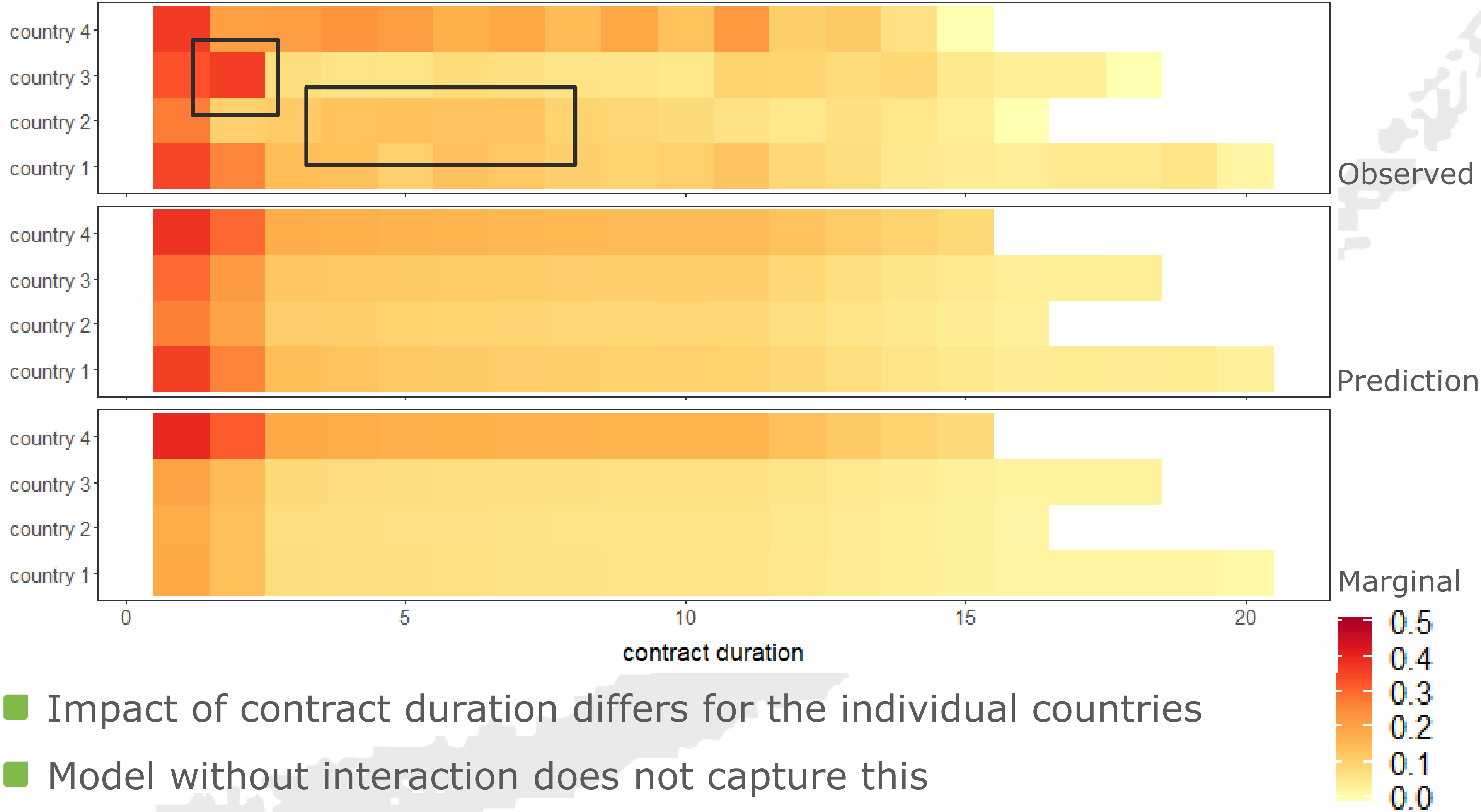
Fused Lasso for entry age





Interactions

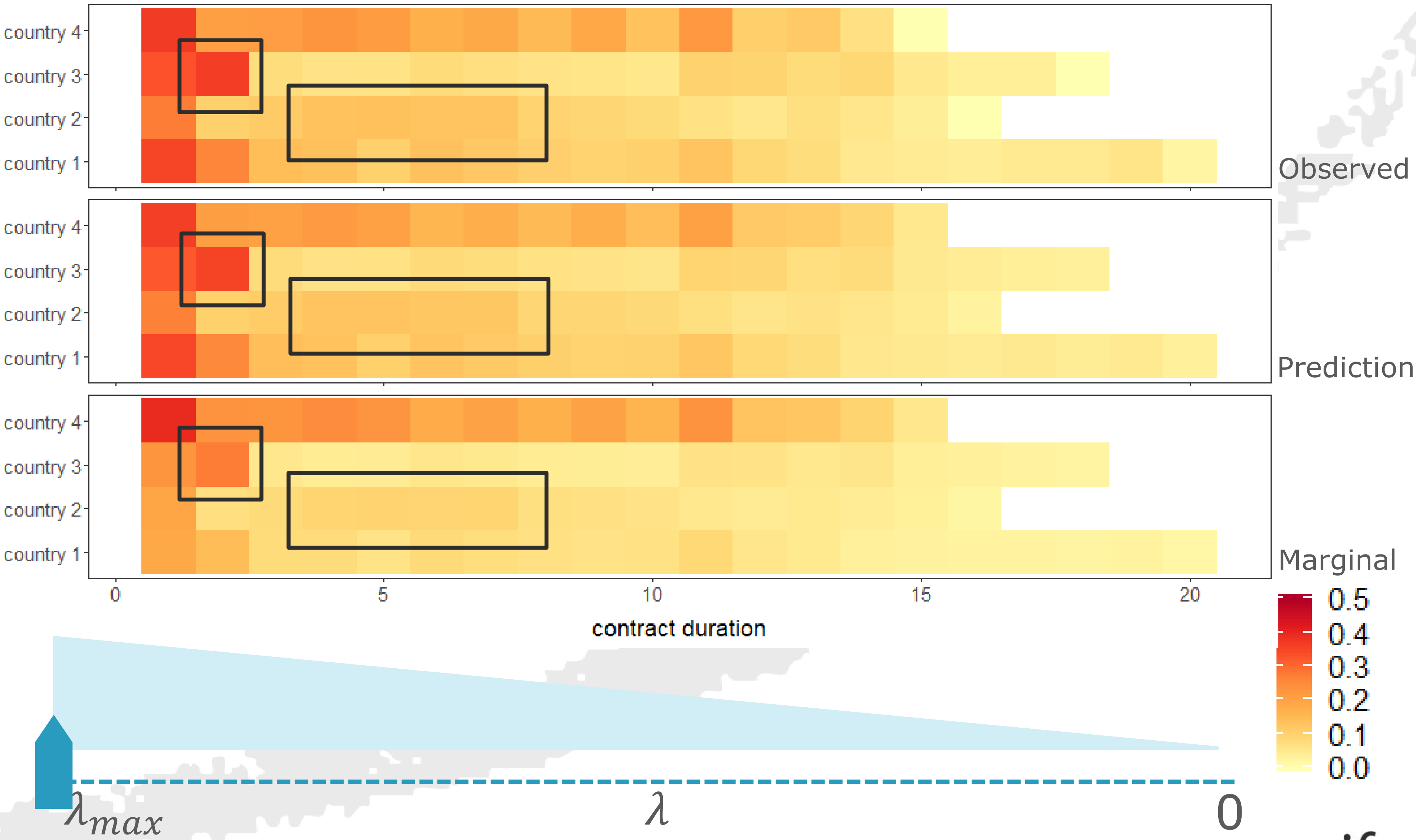
Motivation – Problem of the model without interactions



- 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



Conclusion
Results

Model	Number of parameters	1 - Deviance/ Null Deviance
Intercept Only	1	0%
Whittaker-Henderson	20	6.7%
Lasso without interaction	44 (out of 77)	12.1%
Lasso with interaction	79 (out of 145)	12.9%

+81%

+6%

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

Model	Number of parameters	$1 - D/D_0$
Lasso without interaction	44	12.1%
"Screening" Lasso	30	12.1%
Lasso all regular	70	12.2%
Macroeconomic	72	13.3%
Elastic net, α = 50%	55	12.2%
Offset model	64	12.7%

References

- Eling, M. and Kiesenbauer, D. (2014). What Policy Features Determine Life Insurance Lapse? An Analysis of the German Market. Journal Risk and Insurance 81: 241-269
- Barucci, E., Colozza, T., Marazzina, D. and Rroji, E. (2020). The determinants of lapse rates in the Italian life insurance market. Eur. Actuar. J. 10, 149 - 178
- Tibshirani, R. and Taylor, J. (2011). The solution path of the generalized lasso. The Annals of Statistics 39(3), 1335-1371
- Devriendt, S., Antonio, K., Reynkens, T., & Verbelen, R. (2018). Sparse regression with multi-type regularized feature modeling. Insurance: Mathematics and Economics, 96, 248–261.
- Kiesenbauer, D. (2012). Main Determinants of Lapse in the German Life Insurance Industry. North American Actuarial Journal 16:1, 52–73.
- Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., & Knight, K. (2005). Sparsity and smoothness via the fused lasso. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67: 91–108.

ifa

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