Institut für Finanz- und Aktuarwissenschaften ifa

Wie kann Data Analytics helfen, um Best-Estimate-Schätzungen für das Verhalten von Versicherungsnehmern zu bestimmen?

WIMA Kongress 2022

- Lucas Reck
- 19. November 2022
- Gemeinsame Arbeit mit Andreas Reuß and Johannes Schupp

Agenda

Einleitung

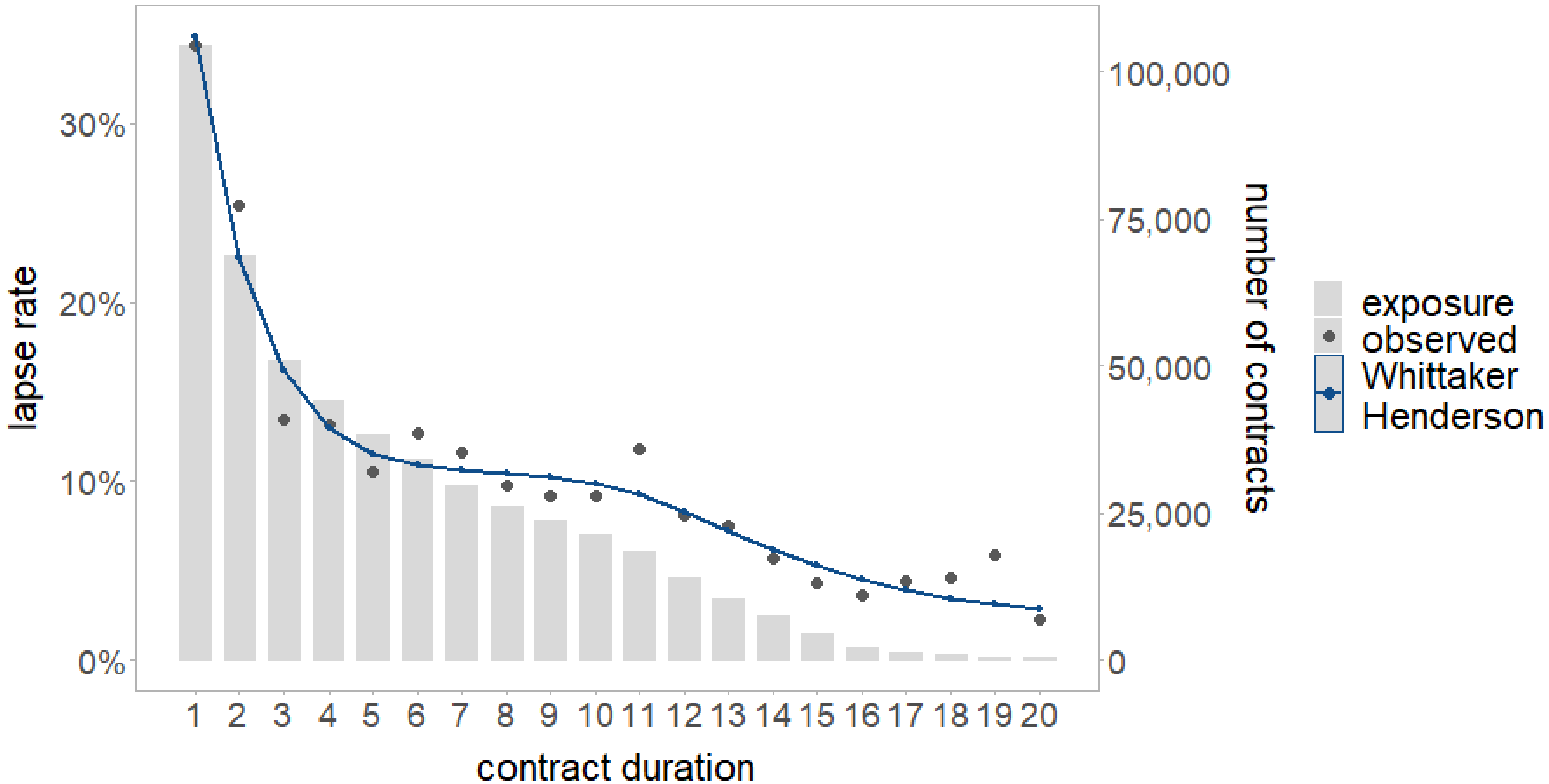
Motivation für ein Stornomodell

- Storno ist eins der Hauptrisiken für Lebensversicherungen.
  - signifikanter Einfluss auf die Zahlungsströme und die Profitabilität von Lebensversicherern
    - relevant für das Asset-Liability-Management und das Liquiditätsrisiko
  - Marktkonsistente Bewertungen basieren auf Best Estimates für künftige Stornoraten.
    - z.B. unter Solvency II (spezielles Risikomodul, welches Storno Risiko adressiert)
- Siehe auch: Reck L., Reuß A., and Schupp J. (2022). Identifying the Determinants of Lapse Rates in Life Insurance: an automated Lasso Approach. European Actuarial Journal.

Einleitung

Vorgehen in der Praxis

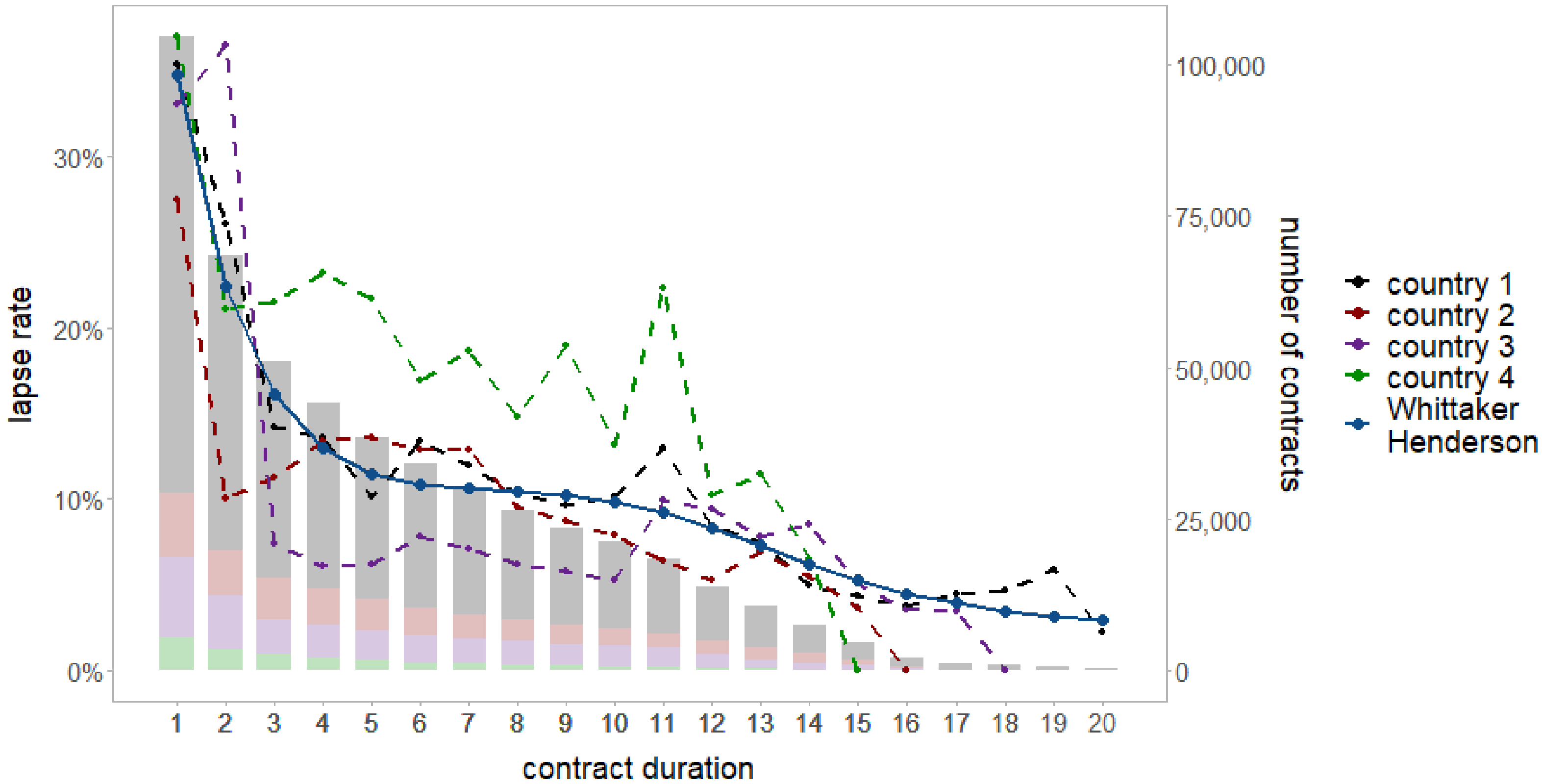
- Whittaker-Henderson (univariater Glättungsalgorithmus)
- Vordefinierte Kovariable (z.B. abgelaufene Vertragsdauer)



Einleitung

Problem am Vorgehen in der Praxis

- Whittaker-Henderson inklusive der Kovariable “Land”



- Das Portfolio wird typischerweise in Teil-Portfolios unterteilt, basierend auf Vertragseigenschaften wie Versicherungstyp, Land, oder Vertriebsweg

Einleitung

Motivation für das Lasso

- Multivariate Modelle – nutzen alle Kovariablen gleichzeitig.
- GLM Storno Model: Eling und Kiesenbauer (2014) und Barucci et al. (2020)
  - Anzahl der Koeffizienten → erheblicher Aufwand
  - Risiko des under- oder overfittings
- Data Analytics Methoden können helfen, beide Probleme zu beheben - wir nutzen einen Lasso Ansatz, um ein Storno Modell abzuleiten, welches
  - automatisch kalibriert und rein datenbasiert ist,
  - komplett interpretierbar bleibt,
  - in der Lage ist, Strukturen innerhalb einer Kovariable zu identifizieren.
- Wir analysieren und kombinieren verschiedene Erweiterungen des Lassos, um die Anforderungen einer praktischen Anwendung zu erfüllen.

Einleitung
Datensatz

- Anwendung
  - Wir verwenden Daten eines europäischen Lebensversicherers, welcher in vier Ländern agiert.
  - Wir nutzen 13 Kovariablen und eine Gesamtanzahl an Beobachtungen von 501.251.
  - Kovariablen sind Standard-Daten einer Versicherung, z.B.:
    - Abgelaufene Vertragsdauer, Eintrittsalter, Versicherungssumme, Land, Versicherungstyp, …

Methodik

Logistische Regression

- Logistische Regression
  - Y_i ist Bernoulli verteilt.
  - $E(Y_i) = p(x_i)$
  - Transformiere $p(x_i)$ und erhalte lineare Abhängigkeit:

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

  - Likelihood Funktion:

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

Methodik

Lasso

- Lasso (Least Absolute Shrinkage and Selection Operator)
  - Ergänze einen Regularisierungsterm:

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

Shrinkage-Faktor: $\lambda \geq 0$
Kontrolliert den Einfluss der Regularisierung und die Modellgüte

Regularisierung:
Bestrafungsterm für die Koeffizienten
Reguläres Lasso: $g(\beta_j) = \sum_{i=1}^{p_j} |\beta_{j,i}|$

Methodik

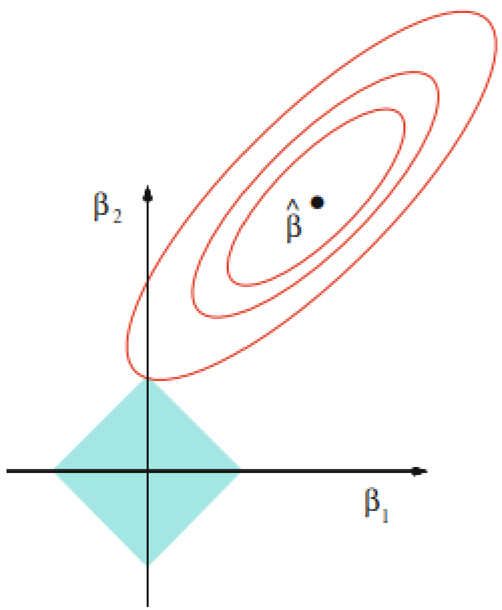
Lasso Veranschaulichung

- Reguläres Lasso mit Regularisierungsterm

$$\min \quad -\log\big(L(\beta, X, y)\big) + \lambda \sum_{i=1}^{m} |\beta_i|$$

- Exemplarisch für zwei Dimensionen:

$$\min \quad -\log\big(L(\beta, X, y)\big) + \lambda\,(|\beta_1| + |\beta_2|)$$



Methodik

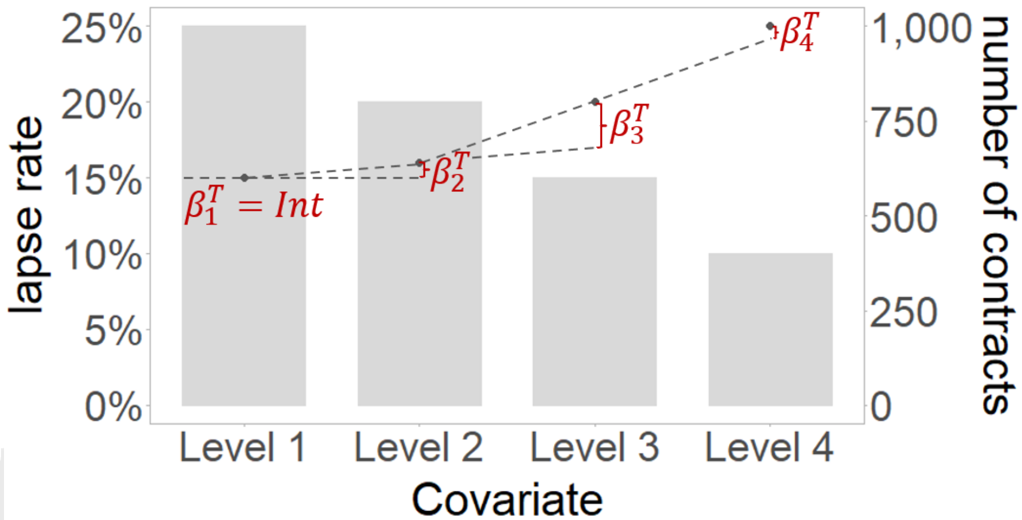
Erweiterung: Fused Lasso und Trend Filtering
Tibshirani and Taylor (2011)

- Jetzt erweitern wir das Lasso: $\min \; -\log\big(L(\beta, X, y)\big) + \lambda \sum_{j=1}^{J} g_j(\beta_j)$
  - Reguläres 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:

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



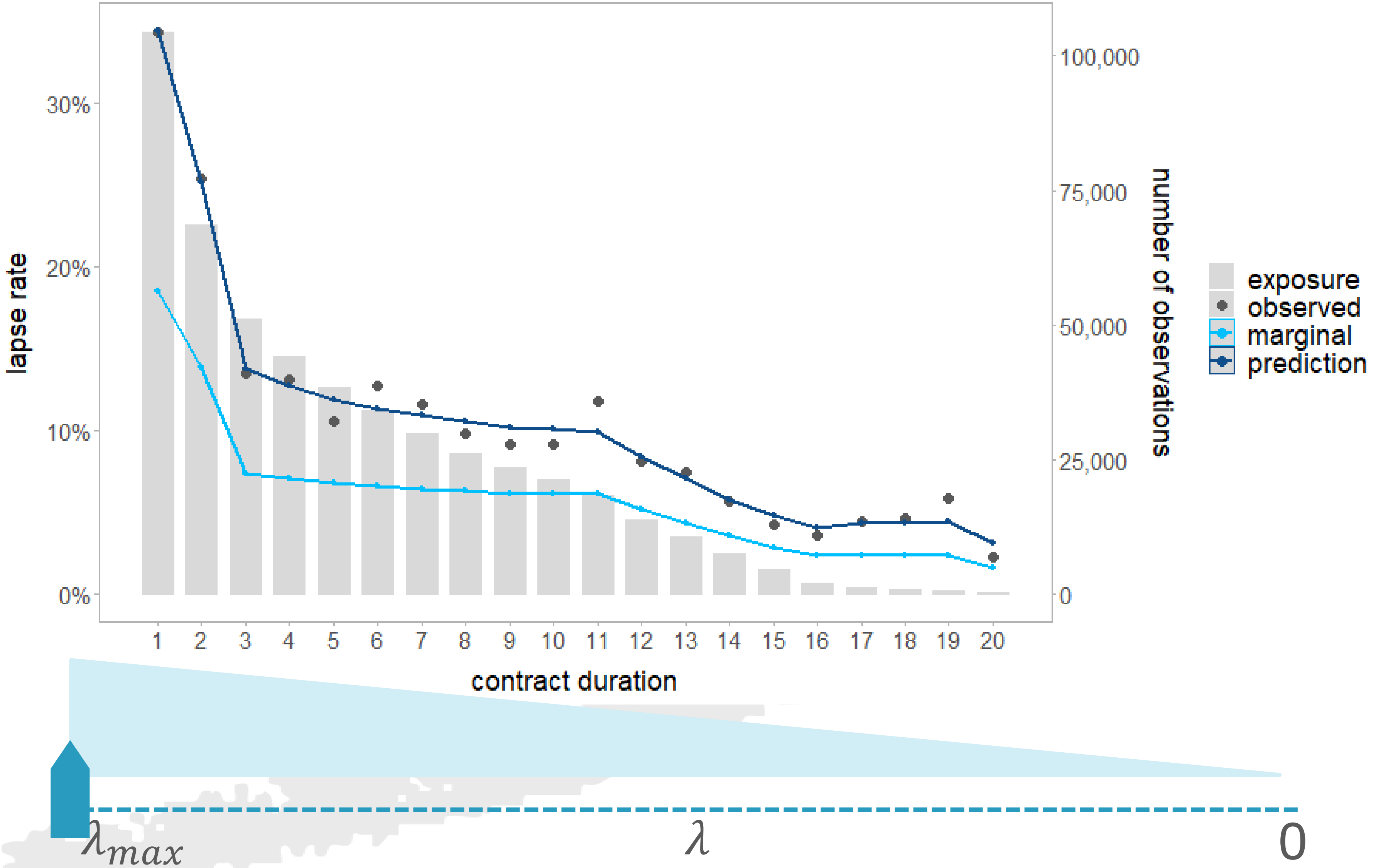
Modell Auswahl

Vorbereitung

- R Schnittstelle für H2O
- Teile jeder Kovariable einen Bestrafungsterm zu:
  - Abgelaufene Vertragsdauer → trend
  - Eintrittsalter → fused
  - Versicherungssumme → trend
  - Land → regulär
  - …
- Hyperparameter λ basiert auf 5-facher Kreuzvalidierung mit der one standard error Regel.
- Devianz als Maß für die Modellgüte

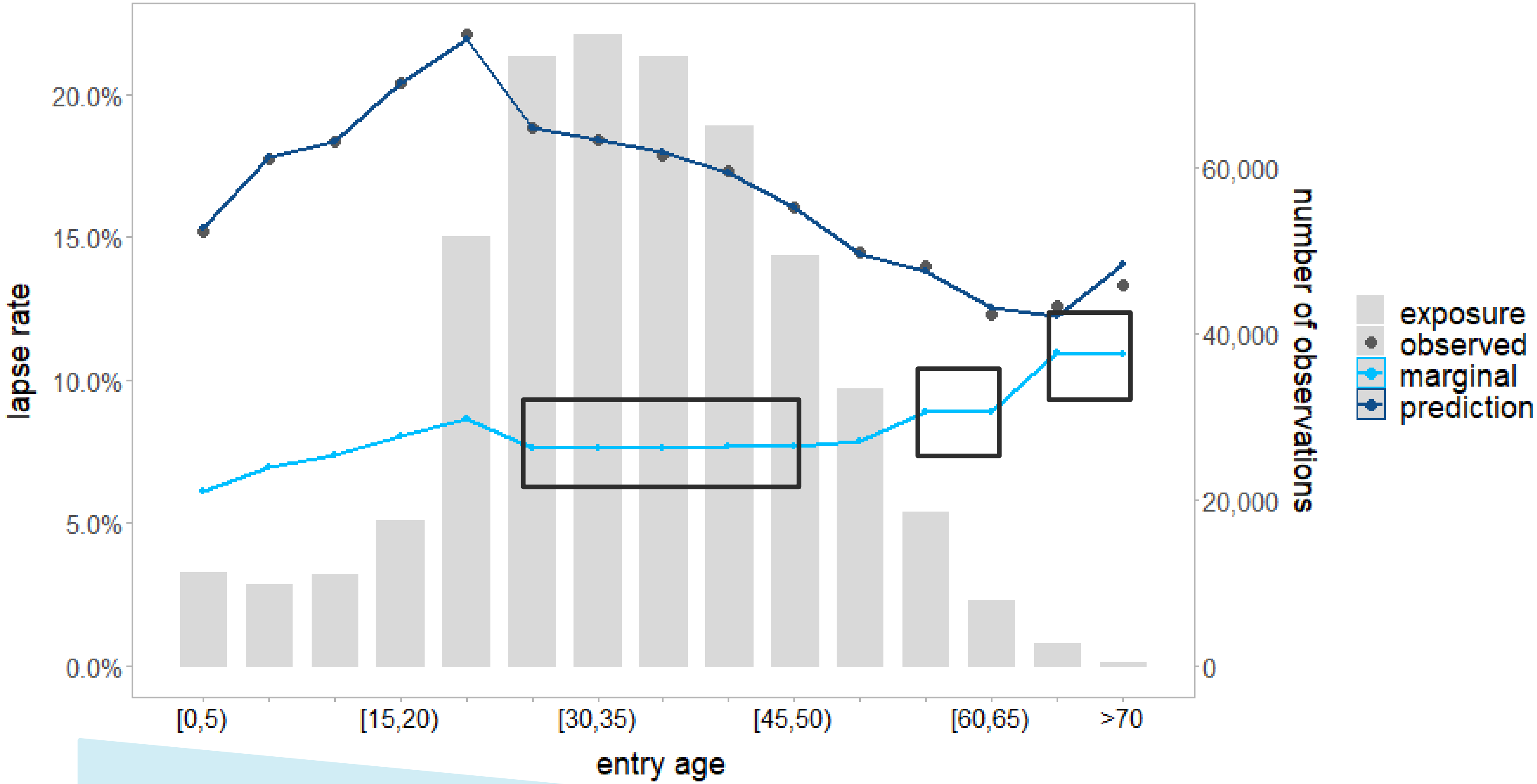
Modell Auswahl

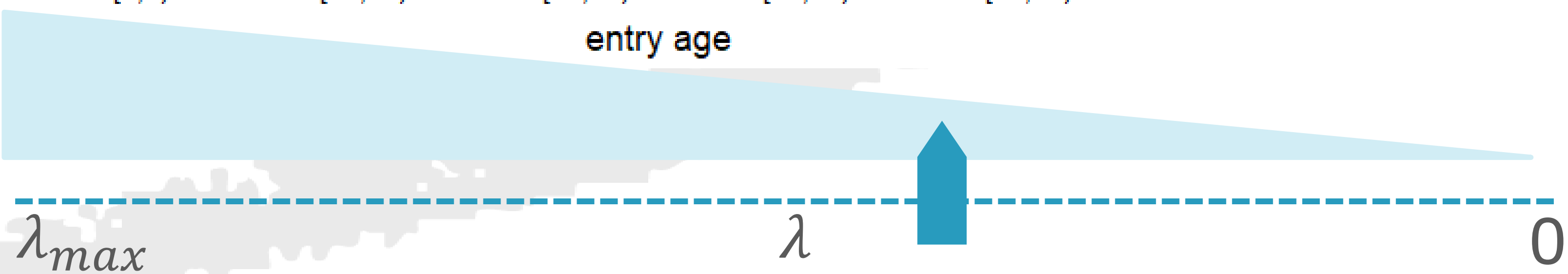
Trend filtering für die abgelaufene Vertragsdauer



Modell Auswahl

Fused Lasso für Eintrittsalter





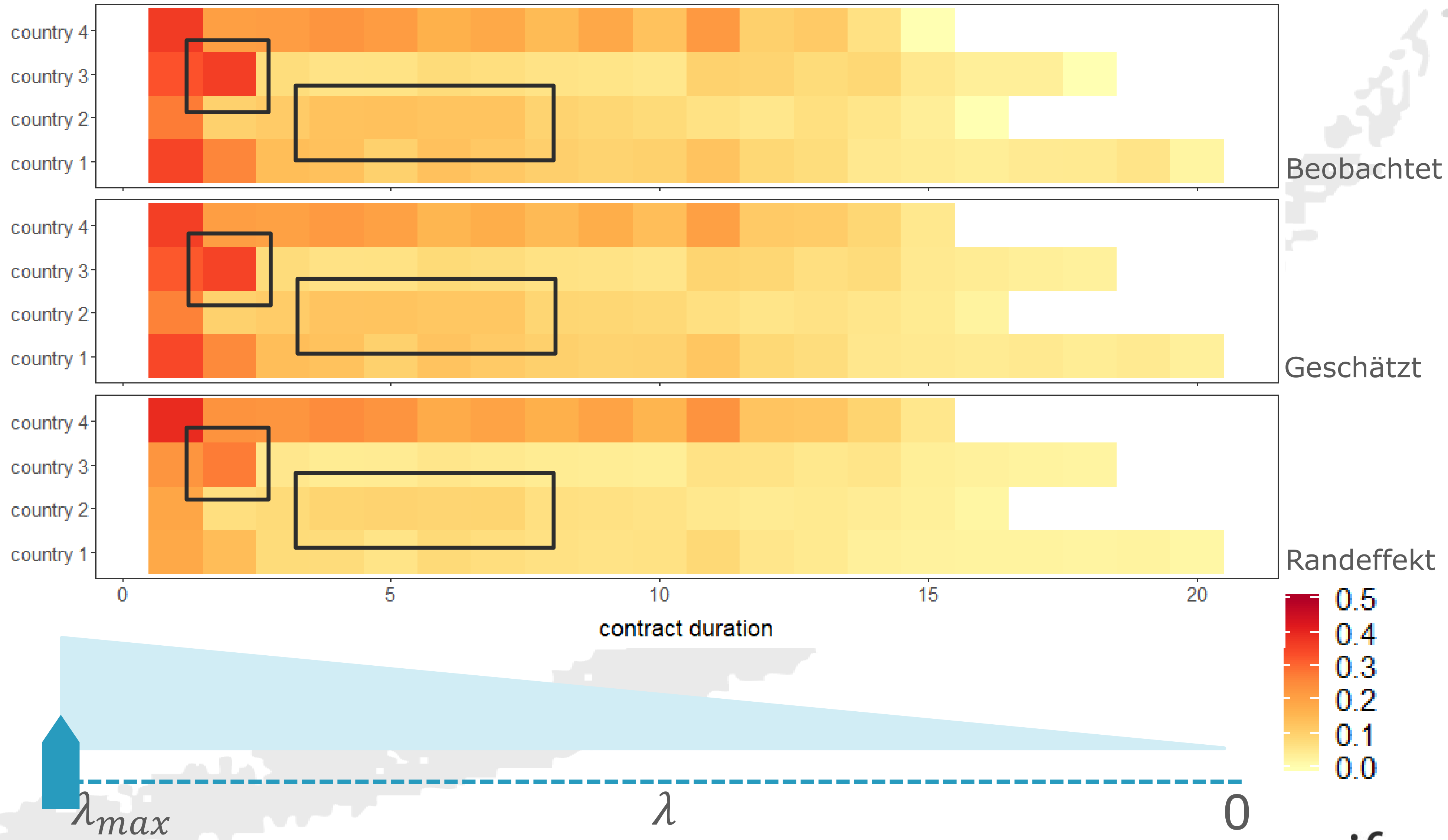
Interaktionen

Motivation – Problem des Modells ohne Interaktionen

- Einfluss der abgelaufenen Vertragsdauer variiert für die einzelnen Länder
- Modell ohne Interaktionen kann dies nicht abbilden
- Wir nehmen die Interaktion abgelaufene Vertragsdauer - Land auf

Interaktionen

Modell mit der Interaktion abgelaufene Vertragsdauer – Land



ifa

Ergebnisse
Bisherige Modelle

Modell	Anzahl Parameter	1 - Devianz/ Null Devianz
Intercept Only	1	0%
Whittaker-Henderson	20	6.7%
Lasso ohne Interaktion	44 (von 77)	12.1%
Lasso mit Interaktion	79 (von 145)	12.9%

+81%

+6%

- Vorteile – Das Modell
  - ist multivariat und nutzt alle Kovariablen gleichzeitig, um die Stornoraten zu schätzen,
  - kalibriert automatisch und ist rein datenbasiert,
  - bleibt komplett interpretierbar,
  - ist in der Lage, Strukturen innerhalb einer Kovariable zu erkennen.

Ergebnisse

Weitere Modelle und Ausblick

- Sensitivitätsanalyse:
  - Basis Modell
  - “Screening” vs “Selecting” Eigenschaft des Lassos
  - Bestrafungstypen
  - Makroökonomische Kovariablen
  - Elastisches Netz
  - Offset Modell für Interaktionen

Modell	Anzahl Parameter	$1 - D/D_0$
Lasso ohne Interaktionen	44	12.1%
“Screening” Lasso	30	12.1%
Lasso alles regulär	70	12.2%
Makroökonomisch	72	13.3%
Elastisch, α = 50%	55	12.2%
Offset Modell	64	12.7%

- Ausblick für zukünftige Forschung
  - Weitere machine learning Ansätze (random forest, neural networks, etc.)
  - Multistate Modell (Aktiv, Beitragsfrei, Storno)

Quellen

- 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
- James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013). An introduction to statistical learning. Springer.
- 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.
- Reck L., Reuß A., and Schupp J. (2022). Identifying the Determinants of Lapse Rates in Life Insurance: an automated Lasso Approach. European Actuarial Journal. https://doi.org/10.1007/s13385-022-00325-1

Kontakt

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