

Mit LASSO Daten einfangen und nutzen

qx-Club Wiesbaden

- Dr. Johannes Schupp
- 09. März 2022



Agenda



Was ist das - LASSO?

Für was braucht man das?

zugrundeliegende Theorie

Wie kann man das einsetzen?

Anwendungsfälle aus dem Alltag

**Institut für Finanz- und
Aktuarwissenschaften**

Für was braucht man das?

Oder: Was der Aktuar eben so macht...

In der S/U Versicherung ermitteln Aktuar auf Basis historischer Beobachtungen **risikoadäquate Prämien**. Diese jährliche Aufgabe bindet viele Ressourcen.

Der Aktuar benötigt Modelle, die Erwartungen an Schäden abhängig von mehreren beobachtbaren Merkmalen möglichst genau bestimmen können.

In der S/U Versicherung sind **neue Datenquellen verfügbar, die Risikomodelle** (und damit Tarifmodelle) **verbessern** können (sozioökonomische Daten, Wasserqualität, Wetterdaten, Telematik).

Der Aktuar benötigt Methoden um aus vielen Daten (Merkmalen) die relevanten Daten zu bestimmen ohne dabei Overfitting zu betreiben.

Im **Risikomanagement einer Lebensversicherung** bestimmen Aktuar auf Basis historischer Beobachtungen Best-Estimate-Annahmen des VN-Verhaltens für künftige Jahre.

- Storno, Beitragsfreistellung, Kapitalwahl, WIK, Dynamiken
- Bestandsprojektion für Solvenzkapital
- Festlegung der strategischen Asset Allocation beim ALM
- Liquiditätsplanung des Unternehmens



Eine Einführung in die zugrundeliegende Theorie

Beginnend bei Bekanntem

Dafür sind Generalized Linear Models (GLM) weit verbreitete und bewährte Modelle. Diese

- + ... sind ein **Supervised Learning Algorithmus**, d.h. eine (beobachtbare) Zielgröße wird auf Basis (beobachtbarer) Merkmale modelliert und für neue beobachtete Merkmale (ohne Zielgröße) angewandt.
- + ... sind **vielseitig**, d.h. verschiedenste Zielgrößen können damit modelliert werden (Schadenfrequenzen, Schadenhöhen, Stornoquoten, Sterblichkeiten, uvm.).
- + ... sind **multivariat**, d.h. zahlreiche Merkmale können auf vielfältige Art gemeinsam modelliert und kombiniert werden.
- + ... sind **flexibel**, da Strukturen in Merkmalen (Trends, Gruppierungen, Polynome, Splines) berücksichtigt werden können.
- ... sind **aufwändig** und **iterativ** zu kalibrieren, da die Modellstrategie für ein Merkmal, alle anderen Risikofaktoren beeinflusst und ggf. ändert.
- ... sind **anfällig** für manuelles **Overfitting**, da man es dabei leicht übertreibt (nicht zu viel / nicht zu wenig Struktur).



Das LASSO behält alle vorteilhaften Eigenschaften eines GLMs. Richtig aufgesetzt, können die Nachteile stark reduziert oder aufgelöst werden.

Eine Einführung in die zugrundeliegende Theorie

Probleme bei der Erstellung eines GLM

- Ein GLM ermöglicht es eine Zielgröße ($\log E(y|x_1, \dots, x_m)$) durch mehrere erklärende Merkmale (z.B. $x_1 = \text{Alter}$, $x_2 = \text{Zahlweise}$, ..., $x_m = \text{Versicherungssumme}$) zu modellieren.

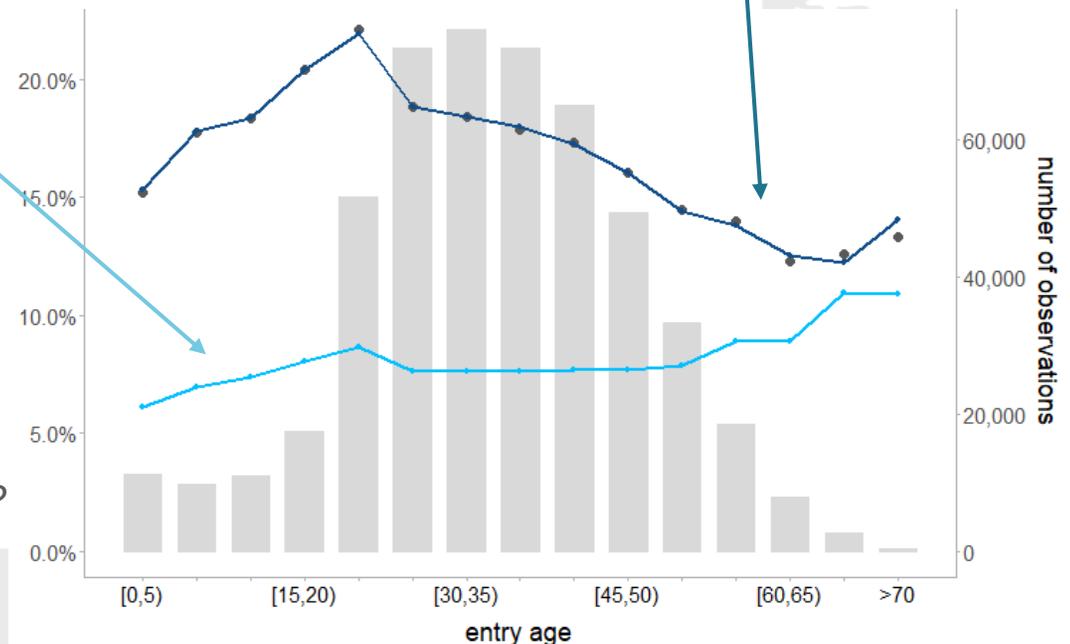
$$\log E(y|x_1, \dots, x_m) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m$$

$$\beta_1^1 x_1^1 + \beta_1^2 x_1^2 + \dots + \beta_1^{p_1} x_1^{p_1}$$

$$\beta_2^1 x_2^1 + \beta_2^2 x_2^2 + \dots + \beta_2^{p_2} x_2^{p_2}$$

$$\beta_m^1 x_m^1 + \beta_m^2 x_m^2 + \dots + \beta_m^{p_m} x_m^{p_m}$$

- Für jedes Merkmal:
welche Strukturen sollen abgebildet werden?
- Welche Merkmale sollen im Modell
berücksichtigt werden?



Beide Fragen können mit LASSO objektiv und automatisiert beantwortet werden.

Eine Einführung in die zugrundeliegende Theorie

Jetzt kommt das LASSO

- Für die Schätzung der Parameter β eines GLMs wird eine Maximum Likelihood Optimierung verwendet.

$$\operatorname{argmin}_{\beta} -\log(L(\beta, X, y))$$

- Diese Schätzung ist in vielen geläufigen Programmen hinterlegt, u.a. SAS, R, Python, Emblem, Matlab ...
- LASSO (Least Absolute Shrinkage and Selection Operator) ist eine Erweiterung des GLMs eingeführt von Tibshirani (1996).

- Erweiterung des GLMs um einen Regularisierungsterm:

$$\operatorname{argmin}_{\beta} -\log(L(\beta, X, y)) + \lambda \sum_{j=1}^m g(\beta_j)$$

Regularisierungsfaktor: $\lambda \geq 0$

Steuerung zwischen Bestrafung und Anpassungsgüte

Bestrafungsart:

Bestrafung der Parameter z.B.:

$$g(\beta_j) = \sum_{i=1}^{p_j} |\beta_{i,j}|$$

Beispiel Zahlart:

$$g(\beta_{ZA}) = \sum_{i=1}^{\omega} |\beta_{i,ZA}|$$

- Durch geschickte Wahl von g können Strukturen und Muster innerhalb von Merkmalen automatisiert erkannt werden.

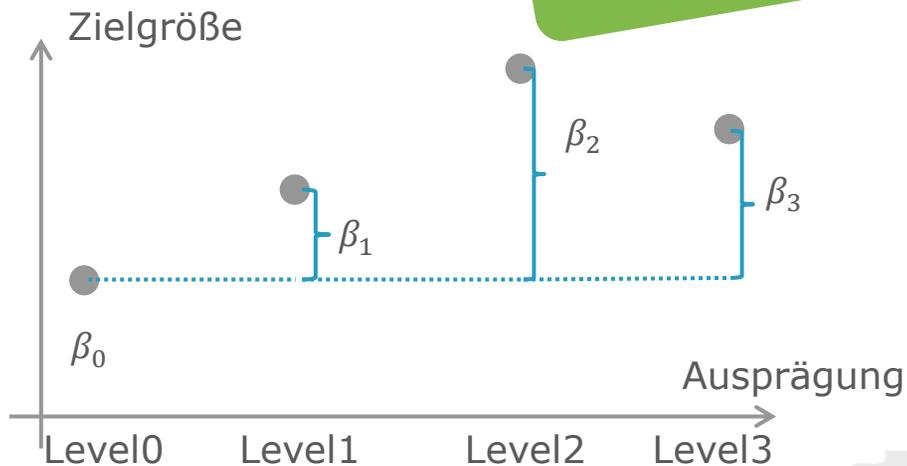
Wie kann man das einsetzen?

Regular LASSO für nominale Merkmale

- Regular LASSO:

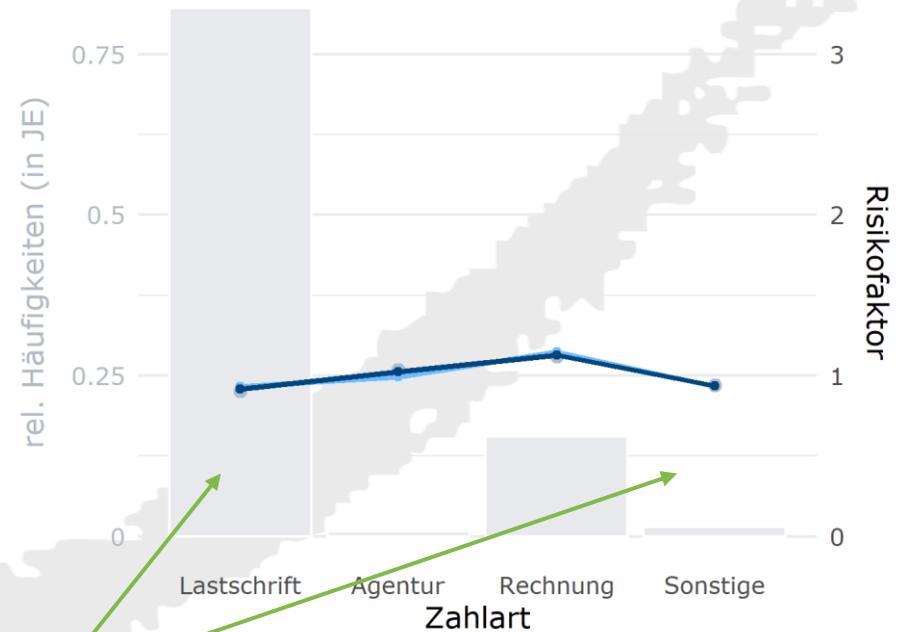
$$g_R(\beta_j) = \sum_{i=1}^{p_j} |\beta_{i,j}|$$

analog GLM



Mit dieser Ausgestaltung beantwortbare Fragen:

- Welche Ausprägungen können über den Intercept modelliert werden und welche nicht?



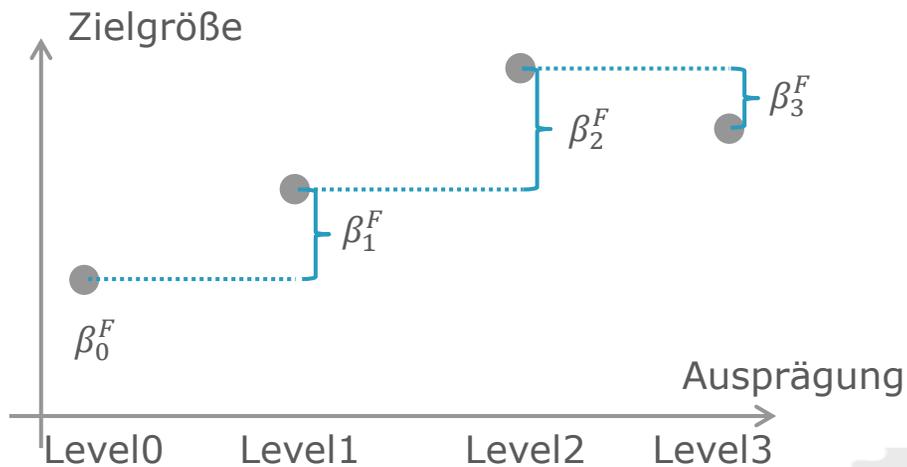
Beispiel Zahlart:
Ausprägung „Sonstige“
kann über Lastschrift
modelliert werden.

Wie kann man das einsetzen?

Fused LASSO für ordinale Merkmale

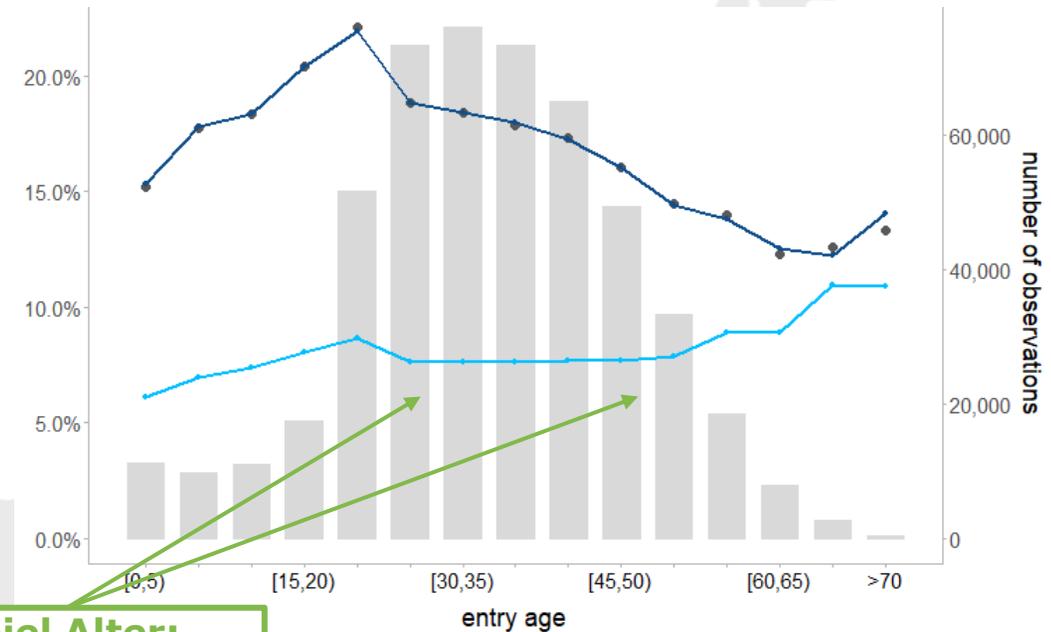
- Fused LASSO:

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



Mit dieser Ausgestaltung beantwortbare Fragen:

- Welche benachbarten Ausprägungen können zusammengefasst werden?



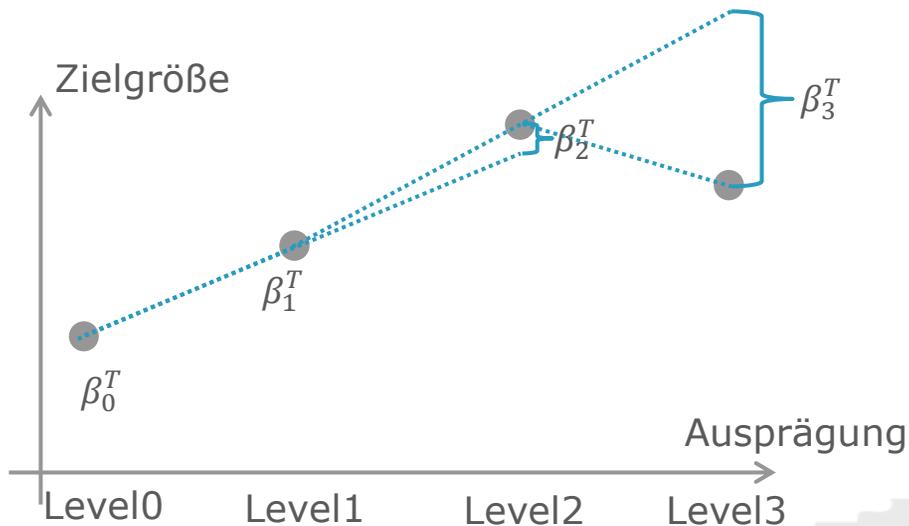
Beispiel Alter:
Gruppierung von
Alter 25-55 zu
einer Altersgruppe

Wie kann man das einsetzen?

Trend filtering LASSO für ordinale Merkmale

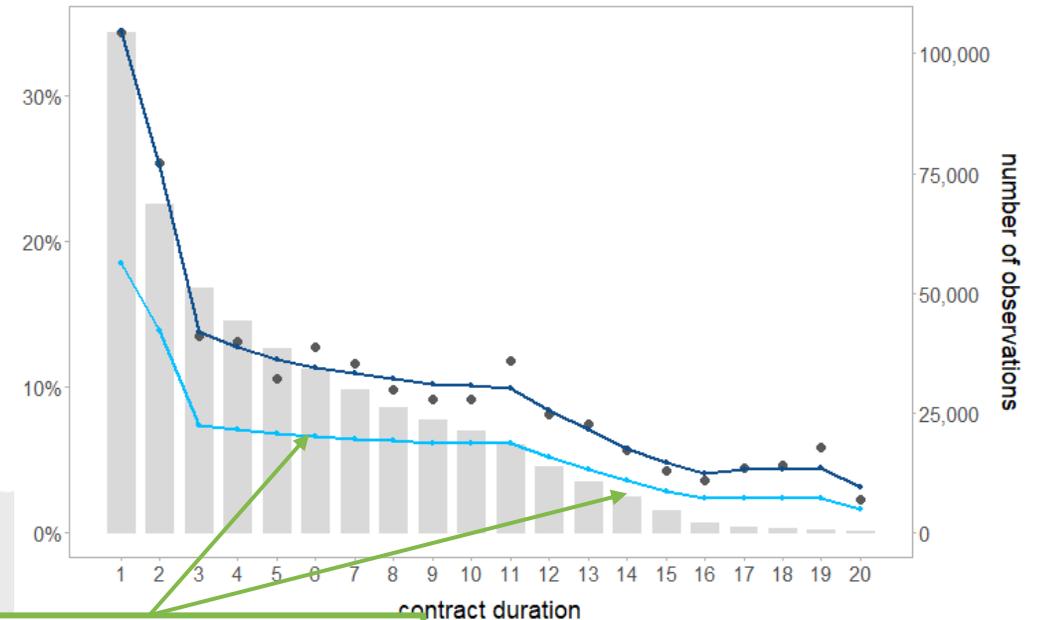
■ Trend Filtering:

$$g_T(\beta_j) = |\beta_{j,1} - \beta_{j,0}| + \sum_{i=3}^{p_j} |\beta_{j,i} - 2\beta_{j,i-1} + \beta_{j,i-2}| =: \sum_{i=1}^{p_j} |\beta_{j,i}^T|$$



Mit dieser Ausgestaltung beantwortbare Fragen:

- Welche benachbarten Ausprägungen können über einen gemeinsamen Trend modelliert werden?



Beispiel Vertragsdauer:
stabile Trends für mehrere Dauern; Zunahme zwischen Dauer 11 und 16

Wie kann man das einsetzen?

- Die drei vorgestellten Arten können kombiniert und erweitert werden.

$$\operatorname{argmin}_{\beta} -\log(L(\beta, X, y)) + \lambda \sum_{j=1}^m g(\beta_j)$$

Regularisierungsfaktor: $\lambda \geq 0$
Steuerung zwischen Bestrafung und Anpassungsgüte

Bestrafungsart:
Bestrafung der Parameter z.B.:
 $g(\beta_j) = \sum_{i=1}^{p_j} |\beta_{i,j}|$

- Nach der Festlung von g für jedes Merkmal verbleibt **EIN** weiterer Freiheitsgrad im Modell.
 - **maximal:** ein sehr hoher **Regularisierungsfaktor** führt dazu, dass $\beta_{i,j} = 0, \forall i, j \rightarrow$ Underfit
 - **minimal:** ein **Regularisierungsfaktor** $\lambda = 0$ führt dazu, dass $\beta_{i,j} \neq 0, \forall i, j \rightarrow$ Overfit
 - **optimal:** ein Ausgleich zwischen Anpassungsgüte und Robustheit gelingt durch Kreuzvalidierung

Einschub Kreuzvalidierung: Die Daten werden z.B. in 5 Teilmengen geteilt und Modelle wiederholt auf unterschiedlichen Teilmengen dieser Daten kalibriert und bewertet.

- Daraus kann abgeleitet werden, welcher Regularisierungsfaktor im Mittel die besten Prognosen liefert.

Agenda



Was ist das - LASSO?

Anwendungsfälle aus dem Alltag

Herleitung von Best-Estimate
Stornoannahmen

Erweiterung der Risikomodellierung durch
viele neue Merkmale

Automatisierung des Tarifierungsprozesses

**Institut für Finanz- und
Aktuarwissenschaften**

Anwendungsfälle aus dem Alltag

Oder: Was der Aktuar eben so macht...

In der S/U Versicherung ermitteln Aktuar auf Basis historischer Beobachtungen **risikoadäquate Prämien**. Diese jährliche Aufgabe bindet viele Ressourcen.

Der Aktuar benötigt Modelle, die Erwartungen an Schäden abhängig von mehreren beobachtbaren Merkmalen möglichst genau bestimmen können.

In der S/U Versicherung sind **neue Datenquellen verfügbar, die Risikomodelle** (und damit Tarifmodelle) **verbessern** können (sozioökonomische Daten, Wasserqualität, Wetterdaten, Telematik).

Der Aktuar benötigt Methoden um aus vielen Daten (Merkmalen) die relevanten Daten zu bestimmen ohne dabei Overfitting zu betreiben.

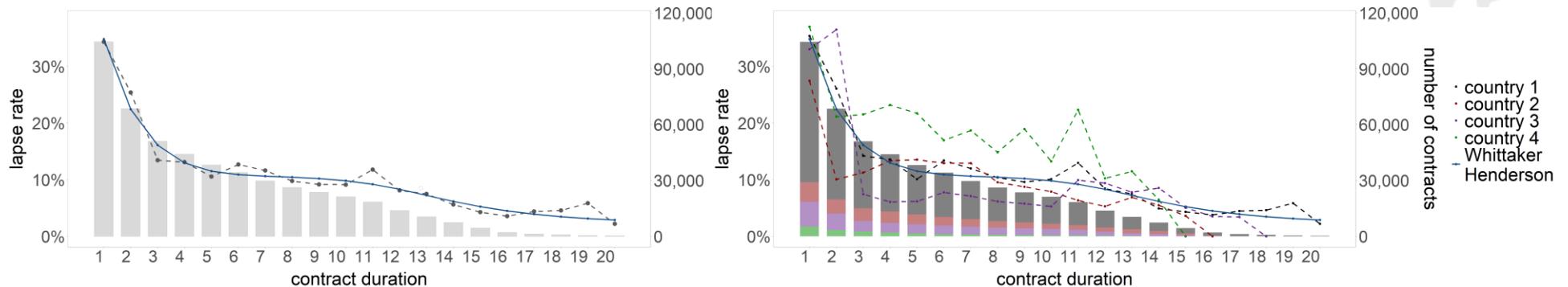
Im **Risikomanagement einer Lebensversicherung** bestimmen Aktuar auf Basis historischer Beobachtungen Best-Estimate-Annahmen des VN-Verhaltens für künftige Jahre.

- Storno, Beitragsfreistellung, Kapitalwahl, WIK, Dynamiken
- Bestandsprojektion für Solvenzkapital
- Festlegung der strategischen Asset Allocation beim ALM
- Liquiditätsplanung des Unternehmens



Herleitung von Best-Estimate Stornoannahmen

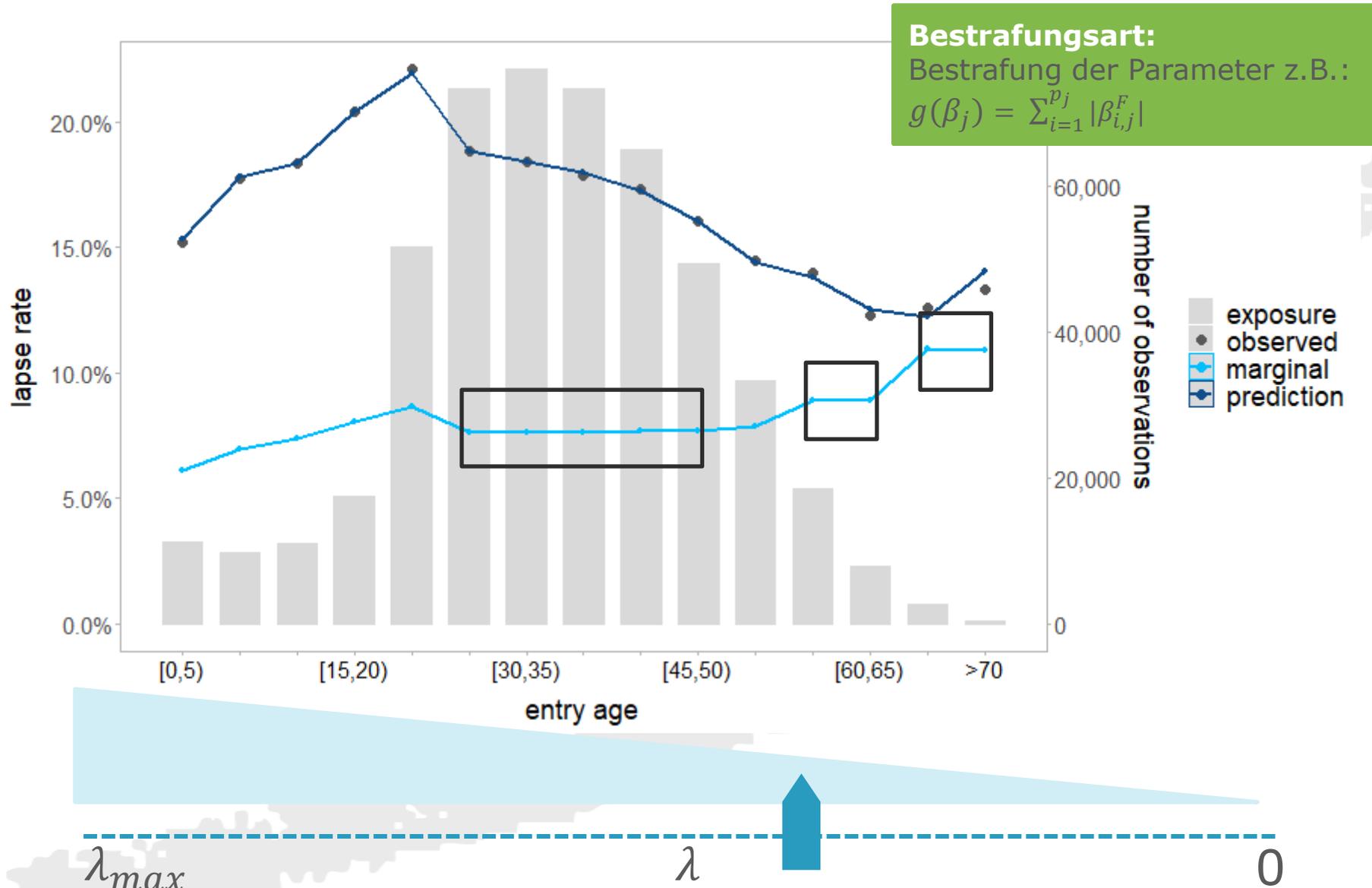
- Im **Risikomanagement** werden Best-Estimate Stornoannahmen benötigt.
- Analyse von **Bestandsdaten eines paneuropäischen Versicherers**



- Diese werden häufig univariat, z.B. mit einer Whittaker-Henderson-Glättung geschätzt.
- Auch hier können multivariate Modelle Ergebnisse verbessern, z.B. in der Form von LASSO.
- Kombination der soeben vorgestellten LASSO-Varianten, um vielfältige Strukturen in Merkmalen abbilden zu können.
 - Regular Lasso: Zielmarkt, Versicherungsart,...
 - Fused Lasso: Alter, Zahlungsperiode,...
 - Trend filtering: abgelaufene Versicherungsdauer, Dynamiksat, Versicherungssumme,...

Herleitung von Best-Estimate Stornoannahmen

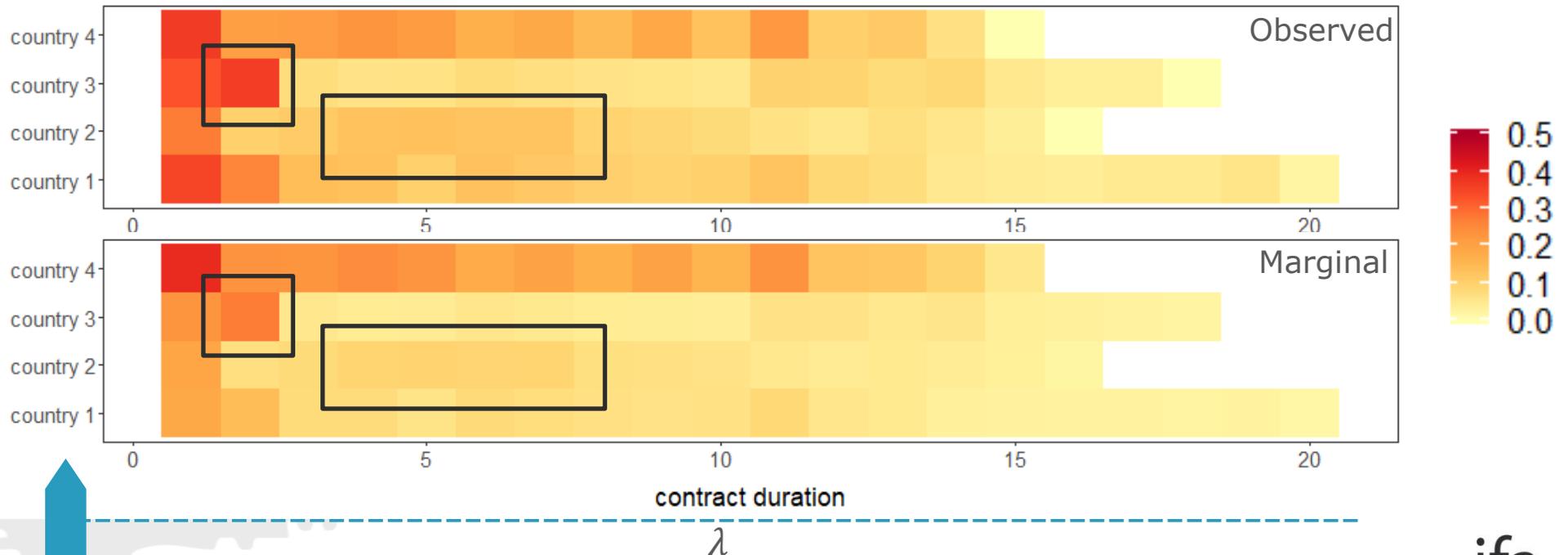
Fused Lasso für Eintrittsalter



Herleitung von Best-Estimate Stornoannahmen

Weitere Ergebnisse

- Multivariate Modelle wie das LASSO sind um etwa **35-40% genauer** als gängige univariate Modelle.
- Dafür sind ausschließlich Daten verwendet worden, die zu den Verträgen ohnehin vorliegen.
- Das LASSO benötigt etwa halb so viele Parameter wie ein GLM und kann automatisiert kalibriert werden - ohne dabei an Performance einzubüßen.
- **Interaktionen** können effizient im Modell berücksichtigt werden.
siehe auch: Reck, L., Reuß, A. und Schupp, J. (2022). Identifying the Determinants of Lapse Rates in Life Insurance: an automated Lasso Approach. Working Paper. Universität Ulm.



Für was braucht man das?

Oder: Was der Aktuar eben so macht...

In der S/U Versicherung ermitteln Aktuar auf Basis historischer Beobachtungen **risikoadäquate Prämien**. Diese jährliche Aufgabe bindet viele Ressourcen.

Der Aktuar benötigt Modelle, die Erwartungen an Schäden abhängig von mehreren beobachtbaren Merkmalen möglichst genau bestimmen können.

In der S/U Versicherung sind **neue Datenquellen verfügbar, die Risikomodelle (und damit Tarifmodelle) verbessern** können (sozioökonomische Daten, Wasserqualität, Wetterdaten, Telematik).

Der Aktuar benötigt Methoden um aus vielen Daten (Merkmalen) die relevanten Daten zu bestimmen ohne dabei Overfitting zu betreiben.

Im **Risikomanagement einer Lebensversicherung** bestimmen Aktuar auf Basis historischer Beobachtungen Best-Estimate-Annahmen des VN-Verhaltens für künftige Jahre.

- Storno, Beitragsfreistellung, Kapitalwahl, WIK, Dynamiken
- Bestandsprojektion für Solvenzkapital
- Festlegung der strategischen Asset Allocation beim ALM
- Liquiditätsplanung des Unternehmens



Erweiterung der Risikomodellierung durch viele neue Merkmale

Ein Beispiel mit sozioökonomischen Daten

In zahlreichen Versicherungsarten können **Schadenquoten** durch Hinzunahme weiterer Merkmale, die bspw. auf PLZ-Ebene/Gemeindeebene vorliegen, **besser erklärt** werden.

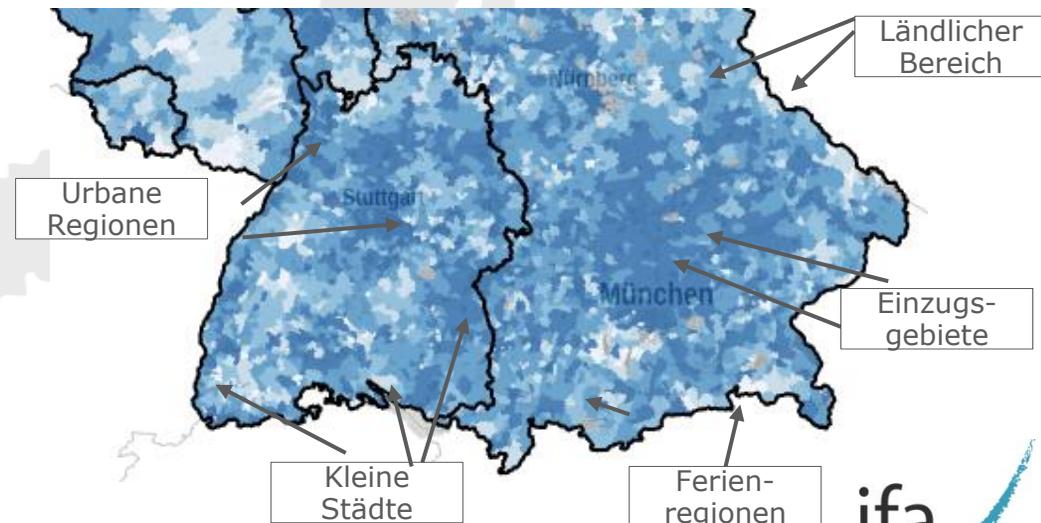
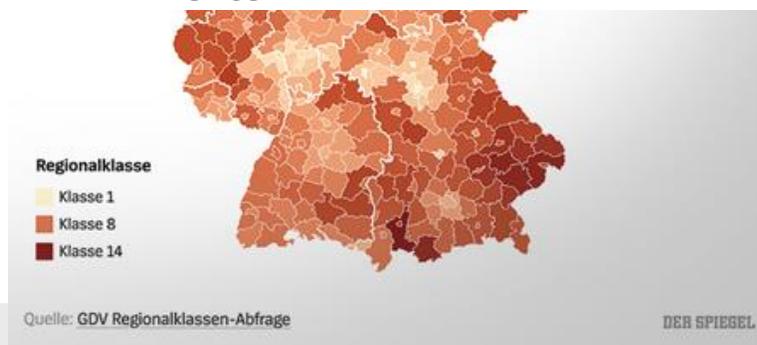
- Beispiel Sozioökonomie: Einbruchdiebstahl in Wohngebäude- oder Hausratversicherung
- Sozioökonomische Daten liegen auf Einzelebene nicht vor und fließen indirekt über korrelierte, z.B. regionale Merkmale ein.
- Um Overfitting zu vermeiden, wird dabei großflächig geglättet (Credibility Ansätze).
- Benachbarte PLZ-Gebiete können sozioökonomisch sehr stark variieren (Blankenese, St. Pauli) oder auch sehr ähnlich sein.
- Differenziertere Betrachtung der sozioökonomischen Effekte über PLZ hinaus erforderlich!



Sinnvoll für die Regionalität!



Nicht sinnvoll für sozioökonomische Effekte!

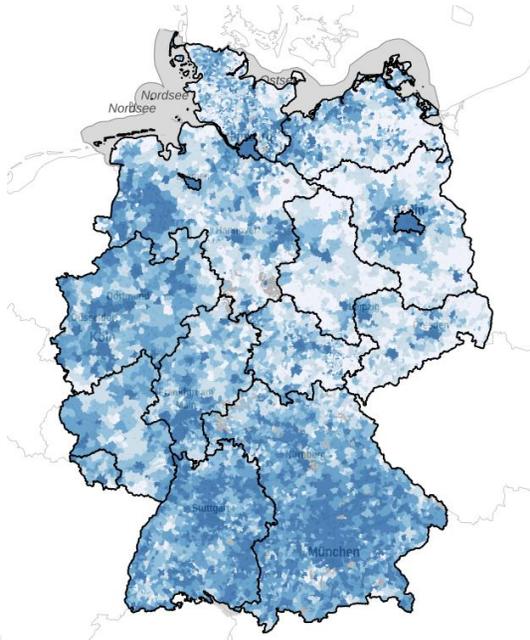


Erweiterung der Risikomodellierung durch viele neue Merkmale

Ein Beispiel mit sozioökonomischen Daten

Daten

- Sozioökonomische Daten können kommerziell erworben werden.
- Es gibt allerdings auch zahlreiche, öffentlich verfügbare Daten.
 - Auch hieraus können bereits mehr als 100 zusätzliche Merkmale verwendet werden.
 - Bsp.: Stadt/Land, Bevölkerungsdichte, Bildungsgrad, Einkommen, Erwerbsquote



Herausforderung:



Qual der Wahl: Welche der vielen Merkmale könnten relevant sein?

Erweiterung der Risikomodellierung durch viele neue Merkmale

Lösungen und Ergebnis

Konkrete Lösung

- LASSO kann automatisiert die Modellbildung und Merkmalselektion übernehmen.
 - gerne auch für 100 + x Merkmale
- Das grundsätzliche Risikomodell (inkl. der Erklärbarkeit) bleibt dabei erhalten.
- Die sozioökonomischen und regionalen Effekte können differenziert betrachtet werden, z.B.:
 - ✓ ein Risikofaktor für die Regionalität
 - ✓ ein Risikofaktor für Sozioökonomie

Projekterfahrung und Ergebnis

Wir haben bereits mehrere Versicherungen bei der Ergänzung ihres klassischen Risikomodells durch moderne Verfahren unterstützt.

Im Bereich Einbruchdiebstahl konnte beispielsweise ein um **10% genaueres Risikomodell** erreicht werden.

Weitere Anwendungsmöglichkeiten

- Vielfältigste Arten von Daten können mit LASSO effizient analysiert werden. Einzige Limitierung: Neue Daten müssen bisherigen Daten zugeordnet werden können.
 - Beispiele: Wetterdaten, geografische Daten, Zahlungsverkehrsdaten, Telematikdaten

Für was braucht man das?

Oder: Was der Aktuar eben so macht...

In der S/U Versicherung ermitteln Aktuar auf Basis historischer Beobachtungen **risikoadäquate Prämien**. Diese jährliche Aufgabe bindet viele Ressourcen.

Der Aktuar benötigt Modelle, die Erwartungen an Schäden abhängig von mehreren beobachtbaren Merkmalen möglichst genau bestimmen können.

In der S/U Versicherung sind **neue Datenquellen verfügbar, die Risikomodelle** (und damit Tarifmodelle) **verbessern** können (sozioökonomische Daten, Wasserqualität, Wetterdaten, Telematik).

Der Aktuar benötigt Methoden um aus vielen Daten (Merkmalen) die relevanten Daten zu bestimmen ohne dabei Overfitting zu betreiben.

Im **Risikomanagement einer Lebensversicherung** bestimmen Aktuar auf Basis historischer Beobachtungen Best-Estimate-Annahmen des VN-Verhaltens für künftige Jahre.

- Storno, Beitragsfreistellung, Kapitalwahl, WIK, Dynamiken
- Bestandsprojektion für Solvenzkapital
- Festlegung der strategischen Asset Allocation beim ALM
- Liquiditätsplanung des Unternehmens



Automatisierung des Tarifierungsprozesses

Gegenüberstellung GLM vs. LASSO

- In vielen Versicherungsarten wird in festen Intervallen, z.B. jährlich ein neues Risikomodell erstellt und darauf aufbauend ein neuer Tarif eingeführt.
 - Die erforderlichen Modellierungsschritte sind identisch und binden enorme Kapazitäten.
- LASSO kann die jährliche Risikomodellerstellung automatisieren und deshalb bereits ab dem zweiten Jahr Aufwände drastisch reduzieren.

	Tarifierung klassisch mit GLM		Tarifierung modern mit LASSO	
Vorgehen	○	iterative und manuelle Modellierung einzelner Merkmale	○	simultane Anpassung aller Merkmale
Interaktionen	⊖	Identifikation von Interaktionen mühsam	⊕	datenbasiertes Erlernen von Strukturen
Aufwand (einmalig)	⊕	kein gesonderter Aufwand	⊖	Aufsetzen des Modells
Aufwand (laufend)	⊖	bei Aktualisierung der Datengrundlage weitgehende Wiederholung aller Schritte	⊕	bei Aktualisierung der Datengrundlage hohe Wiederverwertbarkeit durch automatisierten Prozess
Interpretierbarkeit	⊕	interpretierbare Risikofaktoren	⊕	interpretierbare Risikofaktoren
Robustheit	⊖	anfällig für Overfitting	⊕	robust und objektiv, da rein datenbasiert

Automatisierung des Tarifierungsprozesses

Ein Auszug aus der dafür erforderlichen Implementierung

■ Dafür erforderliche Schritte in R (Python i.W. analog):



■ 1. Datenimport und -aufbereitung:

```
6 library(tidyverse) #Paket für effiziente Datenverarbeitung
7 library(glmnet)   #Paket für LASSO
8 library(MASS)     #Paket für Modellstrategie
9
10 einzeldaten <- read_csv(file="daten_sach.csv",na="-1")
11 einzeldaten %<>% mutate(Zahlart = factor(Zahlart))
```

■ 2. Festlegung einer Modellierungsstrategie je Merkmal und Aufnahme in die Designmatrix

```
14 Kontraste=list(contr.trend(TKL),
15               contr.sdif(Alter_VN),
16               contr.trend(RKL),
17               contr.sdif(Zahlungsperiode),
18               contr.treatment(Zahlart))
19
20 x_design <- model.matrix(Modellgleichung, data=einzeldaten,
21                          contrasts.arg = Kontraste)
22
```

■ 3. Kalibrierung des LASSO

```
26 glm_lasso <- glmnet(x=x_design, y=Schadenanzahl,
27                     family<-"poisson",alpha=1,
28                     standardize = T,weights=JE)
```

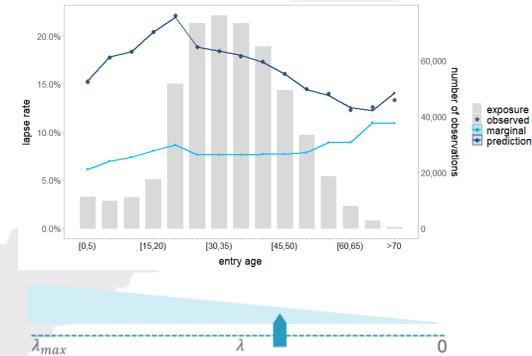
Zusammenfassung

Was der Aktuar mit LASSO eben so macht...

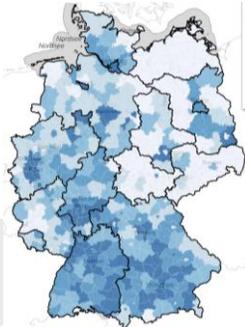
In der S/U Versicherung ermitteln Aktuar auf Basis historischer Beobachtungen **risikoadäquate Prämien**. Diese jährliche Aufgabe bindet viele Ressourcen. LASSO kann automatisiert diese Modelle bilden.



Im **Risikomanagement einer Lebensversicherung** bestimmen Aktuar auf Basis historischer Beobachtungen Best-Estimate Annahmen des VN Verhaltens für künftige Jahre. LASSO ermöglicht bessere und robustere Modelle.



Automatisierbare Verarbeitung immer **neuer Datenquellen**, die Risikomodelle (und damit Tarifmodelle) **verbessern** können



Agenda



Was ist das - LASSO?
Anwendungsfälle aus dem Alltag

**Institut für Finanz- und
Aktuarwissenschaften**

Kontaktdaten

Formale Hinweise

Institut für Finanz- und Aktuarwissenschaften

Kontaktdaten

Dr. Johannes Schupp

+49 (731) 20 644-241

j.schupp@ifa-ulm.de



- Dieses Dokument ist in seiner Gesamtheit zu betrachten, da die isolierte Betrachtung einzelner Abschnitte möglicherweise missverständlich sein kann. Entscheidungen sollten stets nur auf Basis schriftlicher Auskünfte gefällt werden. Es sollten grundsätzlich keine Entscheidungen auf Basis von Versionen dieses Dokuments getroffen werden, welche mit „Draft“ oder „Entwurf“ gekennzeichnet sind. Für Entscheidungen, welche diesen Grundsätzen nicht entsprechen, lehnen wir jede Art der Haftung ab.
- Dieses Dokument basiert auf unseren Marktanalysen und Einschätzungen. Wir haben diese Informationen vor dem Hintergrund unserer Branchenkenntnis und Erfahrung auf Konsistenz hin überprüft. Eine unabhängige Beurteilung bzgl. Vollständigkeit und Korrektheit dieser Information ist jedoch nicht erfolgt. Eine Überprüfung statistischer bzw. Marktdaten sowie mit Quellenangabe gekennzeichnete Informationen erfolgt grundsätzlich nicht. Bitte beachten Sie auch, dass dieses Dokument auf Grundlage derjenigen Informationen erstellt wurde, welche uns zum Zeitpunkt seiner Erstellung zur Verfügung standen. Entwicklungen und Unkorrektheiten, welche erst nach diesem Zeitpunkt eintreten oder offenkundig werden, können nicht berücksichtigt werden. Dies gilt insbesondere auch für Auswirkungen einer möglichen neuen Aufsichtspraxis.
- Unsere Aussagen basieren auf unserer Erfahrung als Aktuare. Soweit wir bei der Erbringung unserer Leistungen im Rahmen Ihrer Beratung Dokumente, Urkunden, Sachverhalte der Rechnungslegung oder steuerrechtliche Regelungen oder medizinische Sachverhalte auslegen müssen, wird dies mit der angemessenen Sorgfalt, die von uns als professionellen Beratern erwartet werden kann, erfolgen. Wenn Sie einen verbindlichen Rat, zum Beispiel für die richtige Auslegung von Dokumenten, Urkunden, Sachverhalten der Rechnungslegung, steuerrechtlichen Regelungen oder medizinischer Sachverhalte wünschen, sollten Sie Ihre Rechtsanwälte, Steuerberater, Wirtschaftsprüfer oder medizinische Experten konsultieren.
- Dieses Dokument wird Ihnen vereinbarungsgemäß nur für die innerbetriebliche Verwendung zur Verfügung gestellt. Die Weitergabe – auch in Auszügen – an Dritte außerhalb Ihrer Organisation sowie jede Form der Veröffentlichung bedarf unserer vorherigen schriftlichen Zustimmung. Wir übernehmen keine Verantwortung für irgendwelche Konsequenzen daraus, dass Dritte auf diese Berichte, Ratschläge, Meinungen, Schreiben oder anderen Informationen vertrauen.
- Jeglicher Verweis auf ifa in Zusammenhang mit diesem Dokument in jeglicher Veröffentlichung oder in verbaler Form bedarf unserer ausdrücklichen schriftlichen Zustimmung. Dies gilt auch für jegliche verbale Informationen oder Ratschläge von uns in Verbindung mit der Präsentation dieses Dokumentes.