

Institut für Finanz- und
Aktuarwissenschaften



DAV/DGVFM
Herbsttagung
2024

Machine Learning zur mehrdimensionalen Modellierung von VN-Verhalten in der LV

Dr. Johannes Schupp (ifa Ulm)
Dr. Lucas Reck (ifa Ulm)

DAV Herbsttagung Mannheim

19.11.2024

Einleitung

Motivation zur Modellierung von VN-Optionen

- Projektion der künftigen Zahlungsströme auf Basis möglichst **realitätsnaher Annahmen** entscheidend für Asset-Liability-Management
- (Einseitiges) Recht der Versicherungsnehmer einer Lebensversicherung, diese Zahlungsströme zu verändern, z.B. durch Storno, (Teil-)Rückkauf, Beitragsanpassungen
- relevant unter Solvency II, z.B. Best Estimate Annahmen für die Ermittlung der vt. Rückstellungen, gestresste Annahmen für die Ermittlung des SCR für Stornorisiken
- Produktfreigabeverfahren (POG): unternehmenseigene Feststellung des angemessenen Kundennutzens (Renditeziele) wenn ein wesentlicher Anteil des Neugeschäfts (50%) storniert hat.
- Es handelt sich dabei also um originäre **aktuarielle Aufgaben**.
- In der Praxis häufig Aufteilung der Daten nach VN-Option und einzelnen Teilbeständen:
 - meist jedoch weitere Informationen verfügbar und relevant → Verbesserung der Prognosegüte möglich

Typischer Ansatz: granulare Teilmodelle

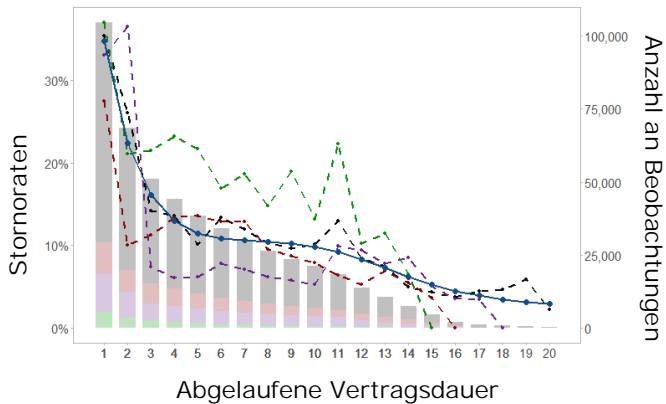
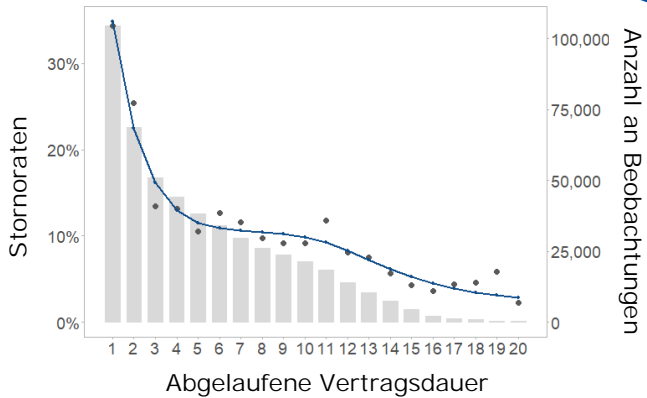
	Klassik	FLV	Riester	BU
• Storno	■	■	■	■
• Beitragsfreistellung	■	■	■	■
• Dynamik	■	■	■	■

Vertragsspezifika

- Alter
- Geschlecht
- Vertragsdauer
- Versicherungssumme

Einleitung

Whittaker-Henderson



Aktuarin: „Wir benötigen ein robustes, einfach interpretierbares und schnell kalibriertes Modell.“

übliches Vorgehen	univariate Glättungsverfahren z.B. Whittaker-Henderson
Verarbeitung vieler Vertragsinformationen	— sehr eingeschränkt
Ausnutzen von Gemeinsamkeiten von Teilbeständen	— nicht möglich
Aufwand für die Erstellung	+ sehr gering
Interpretierbarkeit	+ interpretierbare Einflussfaktoren
Robustheit	+ sehr robust
Prognosegüte	— tendenziell schlecht



Einleitung KI-Ansatz

Data Scientist: „Wir haben tolle neue Datenanalysetools!
Unsere Modelle haben viel mehr Potenzial!“

KI -Ansatz	Beispielsweise neuronale Netze
Verarbeitung vieler Vertragsinformationen	+ möglich
Ausnutzen von Gemeinsamkeiten von Teilbeständen	+ möglich
Aufwand für die Erstellung	- inklusive Fine-Tuning sehr hoch
Interpretierbarkeit	- stark eingeschränkt
Robustheit	- wenig robust
Prognosegüte	+ in der Regel gut



Einleitung

Lasso basierte Modelle

CADS: „Lass uns (gemeinsam) ein Modell entwickeln mit dem Besten aus beiden Perspektiven!“

Lasso-Ansatz

übliches Vorgehen	multivariate Glättungsverfahren Whittaker-Henderson
Verarbeitung vieler Vertragsinformationen	eingeschränkt
Ausnutzen von Gemeinsamkeiten von Teilbeständen	nicht möglich
Aufwand für die Erstellung	sehr gering
Interpretierbarkeit	interpretierbare Einflüsse
Robustheit	sehr robust
Prognosegüte	tendenziell schlecht



Verarbeitung vieler Vertragsinformationen	+
Ausnutzen von Gemeinsamkeiten von Teilbeständen	+
Aufwand für die Erstellung	+
Interpretierbarkeit	+
Robustheit	+
Prognosegüte	+

Key-Features des Lasso

- gleichzeitige Modellkalibrierung und Variablenselektion
- multivariat, performant, datengetrieben, automatisiert und mit guter Prognosegüte
- erkennt vielfältige Strukturen innerhalb der Kovariablen

Beispielsweise neuronale Netze	
Verarbeitung vieler Vertragsinformationen	möglich
Ausnutzen von Gemeinsamkeiten von Teilbeständen	möglich
Aufwand für die Erstellung	sehr gering
Interpretierbarkeit	schlecht
Robustheit	schlecht
Prognosegüte	gut

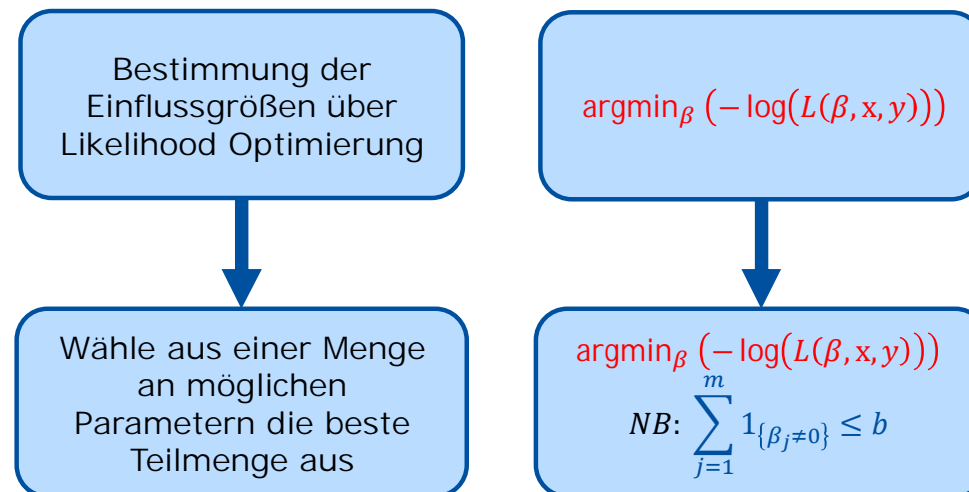
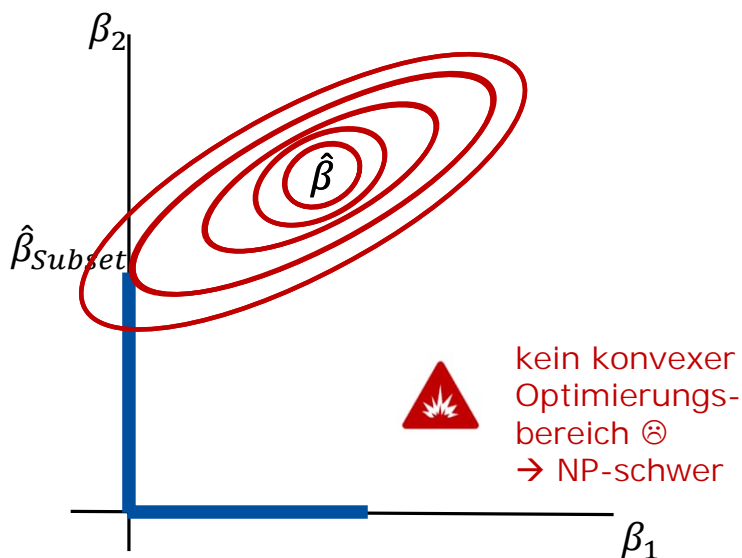


Methodik vom GLM zum Best Subset

Gemeinsame Betrachtung mehrerer Merkmale ($X = X_1, \dots, X_m$) um die Abhängigkeit einer Zielgröße (Y) von interpretierbaren Einflussfaktoren (β) zu modellieren:

$$E(Y|X_1, \dots, X_m) = g^{-1}(X\beta)$$

- Innerhalb jedes Merkmals (z.B. $X_1 = \text{Alter}$) können die Ausprägungen über weitere Strukturen modelliert werden.
 - z.B. Auswahl von Altersgruppen, linearer Trend für Versicherungssummen, ...
 - Ohne weiteres Vorwissen oder Vorverarbeitung sind das potenziell sehr viele Parameter.

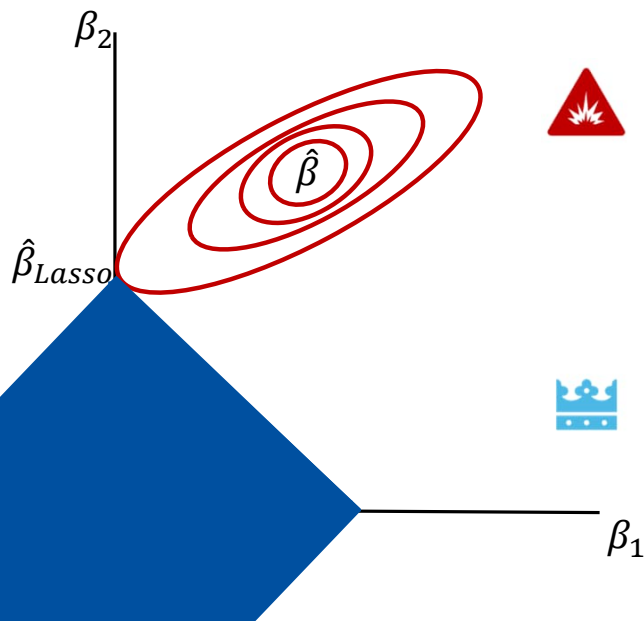


Methodik vom Best Subset zum Lasso

Lasso (Least Absolute Shrinkage and Selection Operator)

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

- L_1 -Bestrafung der Summe der **absoluten** Parameterschätzer (ohne Intercept) und Regularisierungsfaktor λ



kein konvexer
Optimierungs
bereich ☹
→ NP-schwer



geradeso
konvexer
Raum ☺
→ Numerik

Wähle aus einer Menge
an möglichen
Parametern die beste
Teilmenge aus

Wähle aus einer Menge
an möglichen
Parametern die beste
Teilmenge aus

$$\operatorname{argmin}_{\beta} \left(-\log(L(\beta, x, y)) \right)$$

NB: $\sum_{j=1}^m 1_{\{\beta_j \neq 0\}} \leq b$

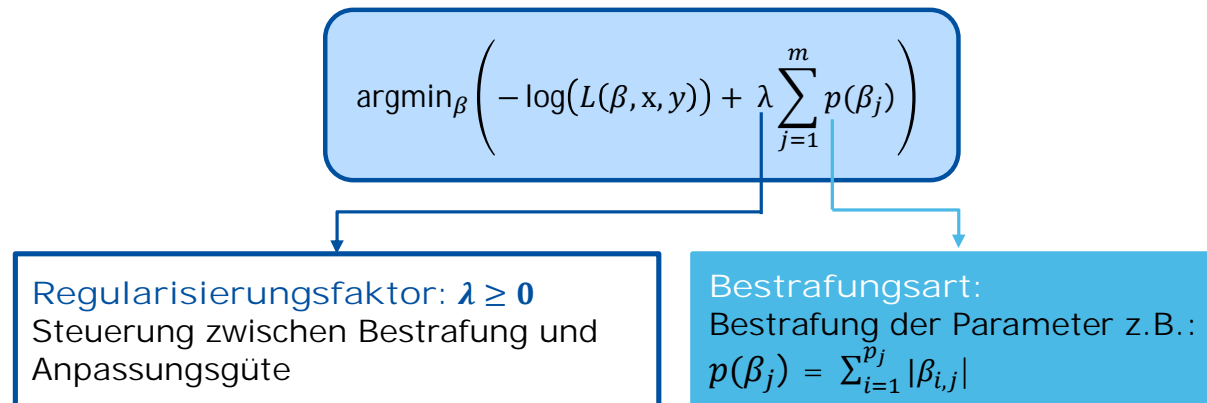
$$\operatorname{argmin}_{\beta} \left(-\log(L(\beta, x, y)) \right)$$

NB: $\sum_j |\beta_j| \leq t$

Methodik

Erweitertes Lasso

Lasso mit unterschiedlichen Bestrafungsarten



- Idee: Man gibt sehr viele Merkmale und Strukturen hinein und das Modell wählt hieraus automatisch die Besten aus. Somit wird nur ein Bruchteil der Einflussgrößen verwendet, ohne signifikant an Prognosegüte zu verlieren.
 - Durch geschickte Wahl von p können verschiedene Strukturen und Muster innerhalb von Merkmalen automatisiert erkannt werden (Designentscheidung).
- Ergebnisse und mögliche Designentscheidungen werden anhand eines Datensatzes eines pan-europäischen Lebensversicherers veranschaulicht.

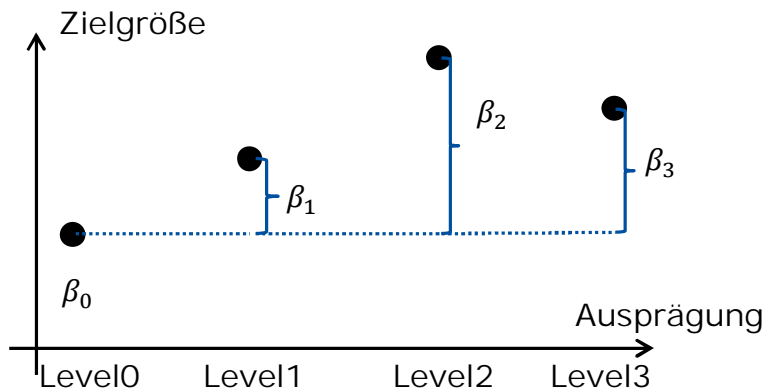
Ergebnisse

Regular Lasso für nominale Merkmale

Regular Lasso

Methodik

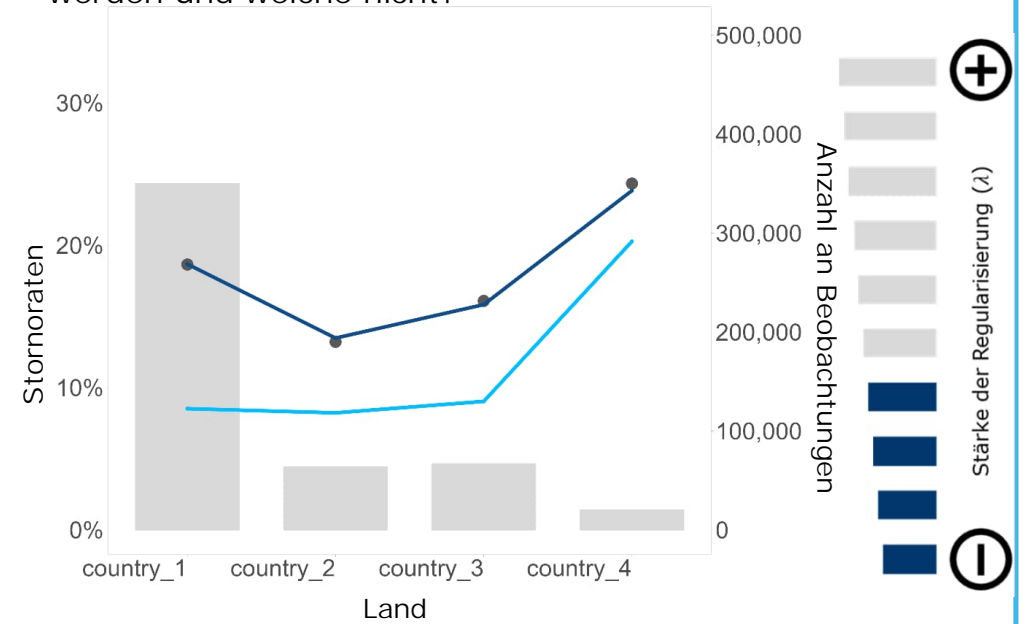
$$p_R(\beta_j) = \sum_{i=1}^{p_j} |\beta_{i,j}|$$



Mit dieser Ausgestaltung beantwortbare Frage:

Ergebnisse

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



- Anzahl
- Beobachtet
- Randeffect
- Vorhersage

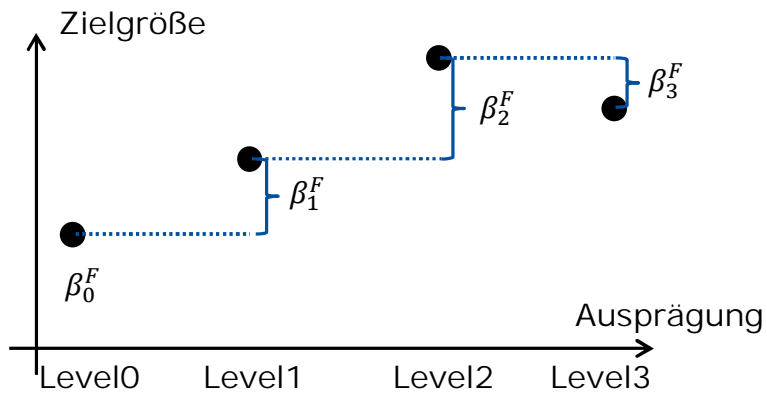
Ergebnisse

Fused Lasso für ordinale Merkmale

Fused Lasso

Methodik

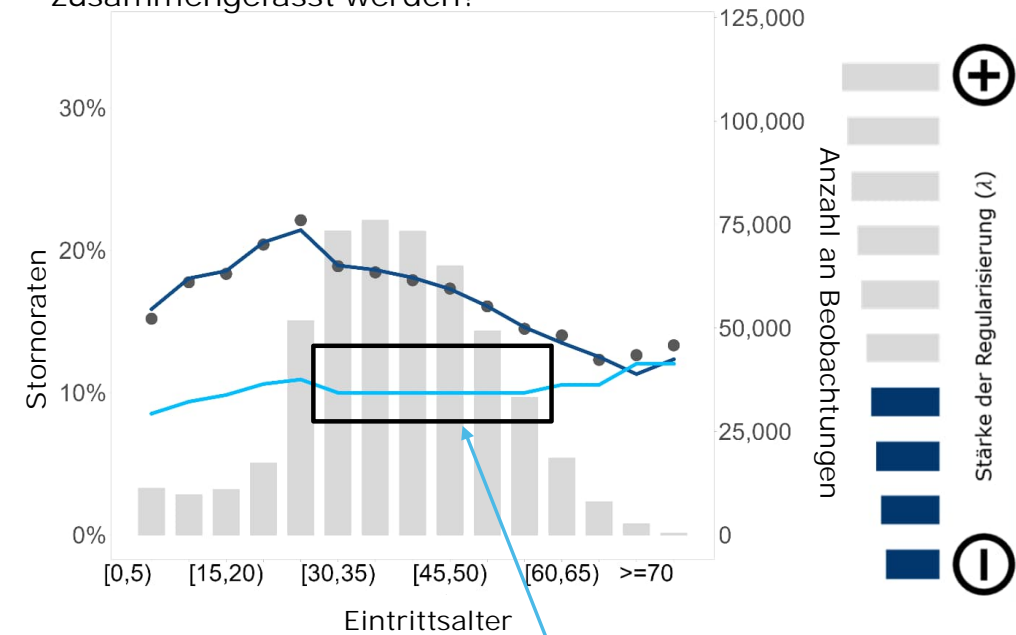
$$p_F(\beta_j) = |\beta_{j,1}| + \sum_{i=2}^{p_j} |\beta_{j,i} - \beta_{j,i-1}| =: \sum_{i=1}^{p_j} |\beta_{j,i}^F|$$



Mit dieser Ausgestaltung beantwortbare Frage:

Ergebnisse

- Welche benachbarten Ausprägungen können zusammengefasst werden?



Anzahl
 Beobachtet
 Randeffekt
 Vorhersage

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

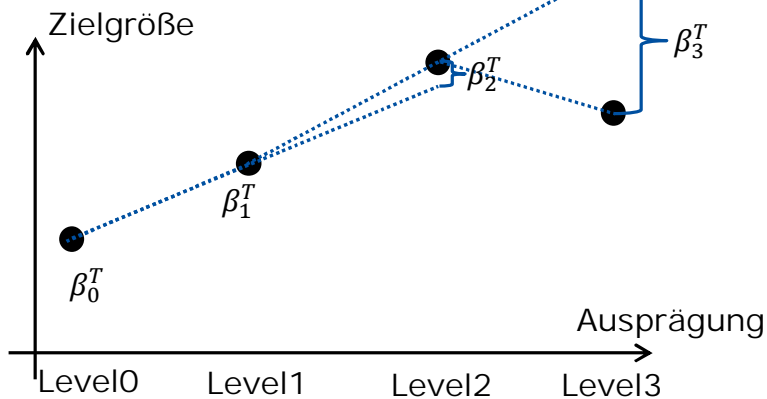
Ergebnisse

Trend filtering für ordinale Merkmale

Trend filtering

Methodik

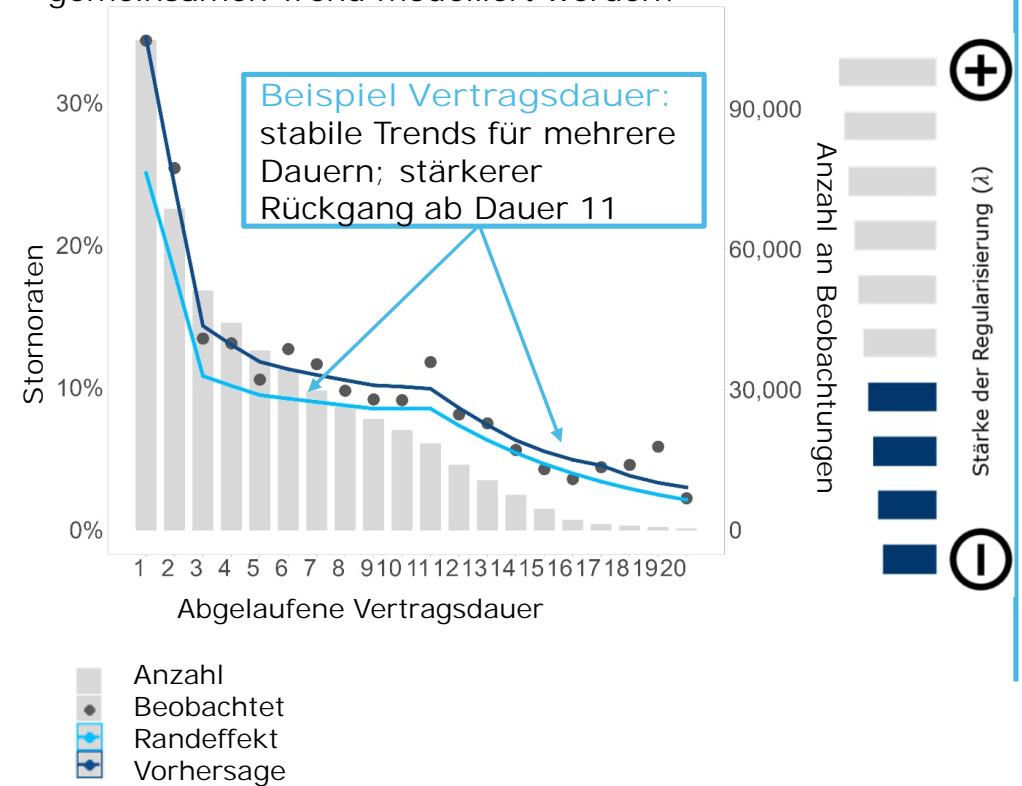
$$p_T(\beta_j) = |\beta_{j,1}| + |\beta_{j,2} - 2\beta_{j,1}| + \sum_{i=3}^{p_j} |\beta_{j,i} - 2\beta_{j,i-1} + \beta_{j,i-2}| =: \sum_{i=1}^{p_j} |\beta_{j,i}^T|$$



Mit dieser Ausgestaltung beantwortbare Frage:

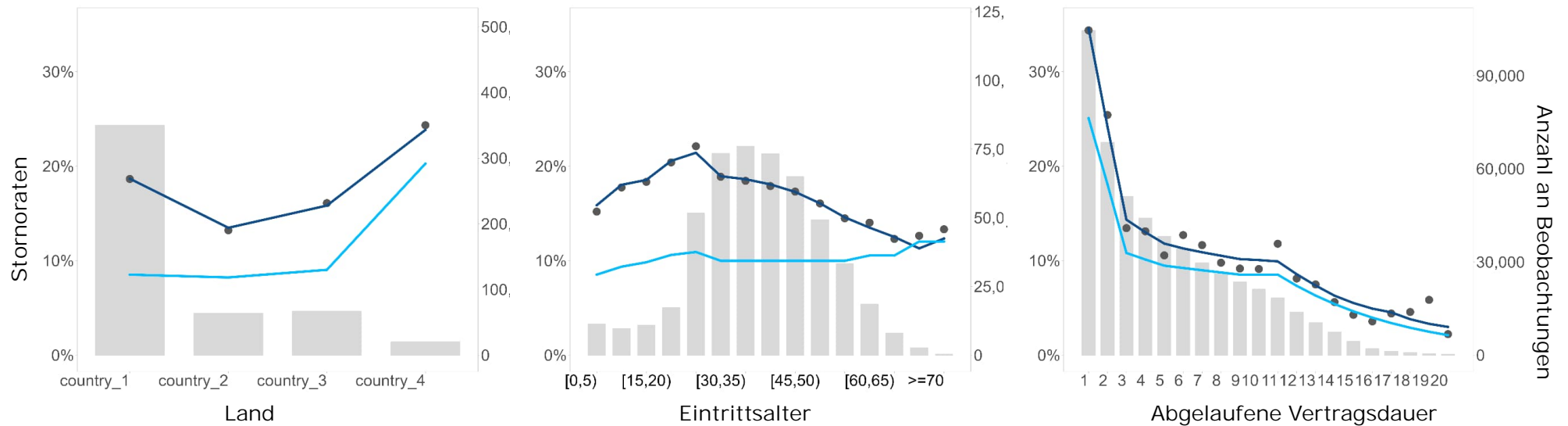
Ergebnisse

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



Ergebnisse

EIN Hyperparameter zur Steuerung der Anpassung



Anzahl
 Beobachtet
 Randeffect
 Vorhersage



Quelle: Reck, L., Schupp, J., & Reuß, A. (2023). Identifying the determinants of lapse rates in life insurance: an automated Lasso approach. *European Actuarial Journal*, 13(2), 541-569.

Weitere Konfigurationsmöglichkeiten und Anwendungen Designentscheidungen und Fallstricke

Ausreißer beim Randeffekt

- große Trendänderungen
- Sprünge

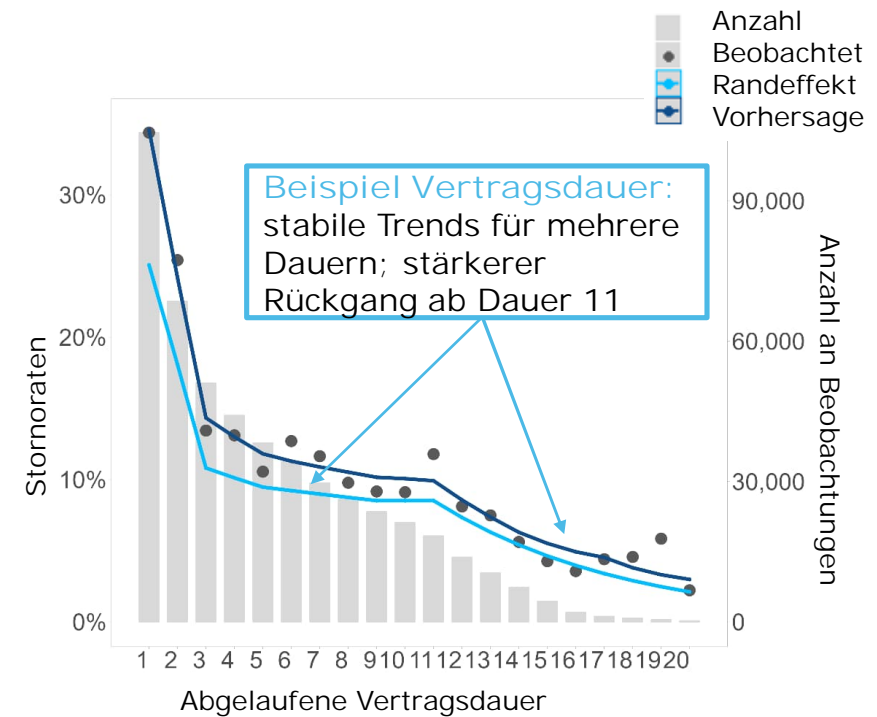
- Idee: Man gibt sehr viele Strukturen / Variationsmöglichkeiten vor und die Optimierung findet die für die Prognose wirklich relevanten.

- Regularisierungsterme ($\lambda \sum_{j=1}^m p(\beta_j)$) erweitert zu:

$$p_T(\beta_{\text{abgelaufene Vertragsdauer}}) + p_F(\beta_{\text{abgelaufene Vertragsdauer}})$$

Trend Filtering
Findet den zugrundeliegenden Trend

Fused Lasso
Findet den Sprung / Ausreißer



Quelle: Reck, L. (2024). The Automation of Core Actuarial Modelling Tasks - an Analysis and Evaluation of the Lasso. Working Paper, Ulm University.

Weitere Konfigurationsmöglichkeiten und Anwendungen Designentscheidungen und Fallstricke

Ausreißer beim Randeffekt

- große Trendänderungen
- Sprünge

Wahl der Kovariablen

- Welche Informationen liegen in welcher Qualität vor?
- Welche weiteren (ggf. externen) Informationen können/sollen ergänzt werden?

Binning

- Welche Ausprägungen können/sollen vorab zusammengefasst werden?
- Wie wählt man sinnvolle Bins?

Bestrafungsterm für Interaktionen

- Kovariable nun 2-dimensional
- Was zählt als „benachbarte“ Ausprägung?

Wahl des Hyperparameters λ

- statistisch motiviert mit Kreuzvalidierung
- anwendungsorientiert
- „Screening vs. Selection“

Startposition beim trend filtering,
...

Quelle: Reck, L. (2024). The Automation of Core Actuarial Modelling Tasks - an Analysis and Evaluation of the Lasso. Working Paper, Ulm University.



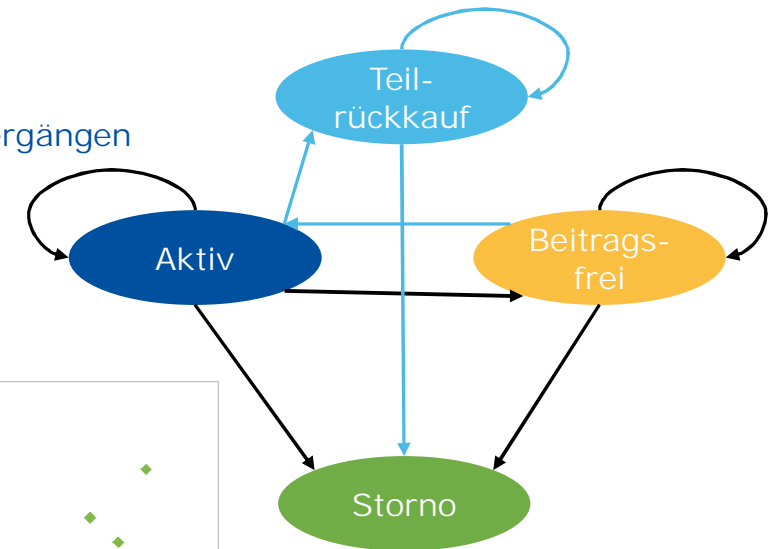
Das Aufsetzen eines verlässlichen Lasso-Modells geht **nicht auf „Knopfdruck“** – um das volle Potenzial von Lasso zu nutzen muss man sowohl die Details des Modellierungsansatzes als auch den Anwendungskontext genau kennen.

Weitere Konfigurationsmöglichkeiten und Anwendungen

Mehrere VN-Optionen - Motivation

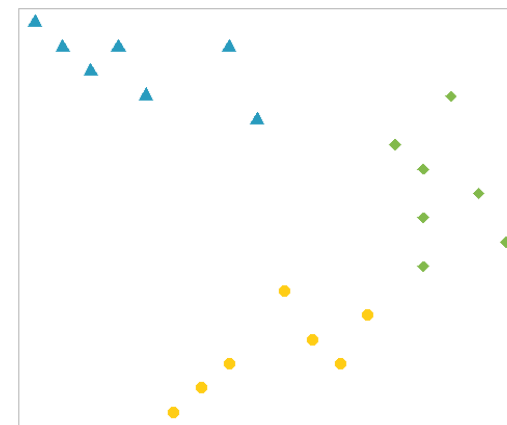
In der Praxis: Ausübung **mehrerer VN-Optionen** mit entsprechenden **Zustandsübergängen**

- Welche Zustände und welche Übergänge sind relevant (materiell)?
- Beispielsweise neben „Storno“ auch Beitragsfreistellung
 - potentiell aber auch weitere Zustände wie Teiltrückkauf
 - oder weitere Übergänge wie Wiederinkraftsetzung



Typischer Ansatz: granulare Teilmodelle

	Klassik	FLV	Riester	BU
• Storno	 	 	 	
• Beitragsfreistellung	 	 	 	
• Dynamik	 	 	 	



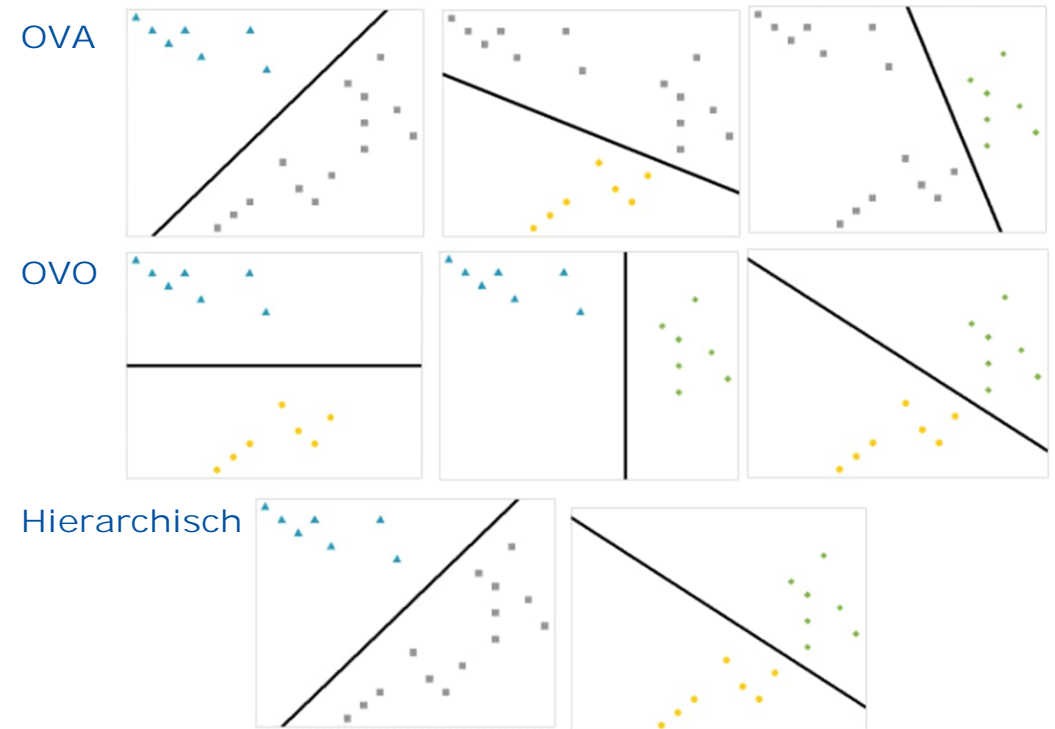
▲ class A ● class B ◆ class C

- Konsistenz der Teilmodelle wichtig:
 - z.B. Summe der Wechselwahrscheinlichkeiten = 1 durch granulare Teilmodelle i.d.R. verletzt.

Weitere Konfigurationsmöglichkeiten und Anwendungen

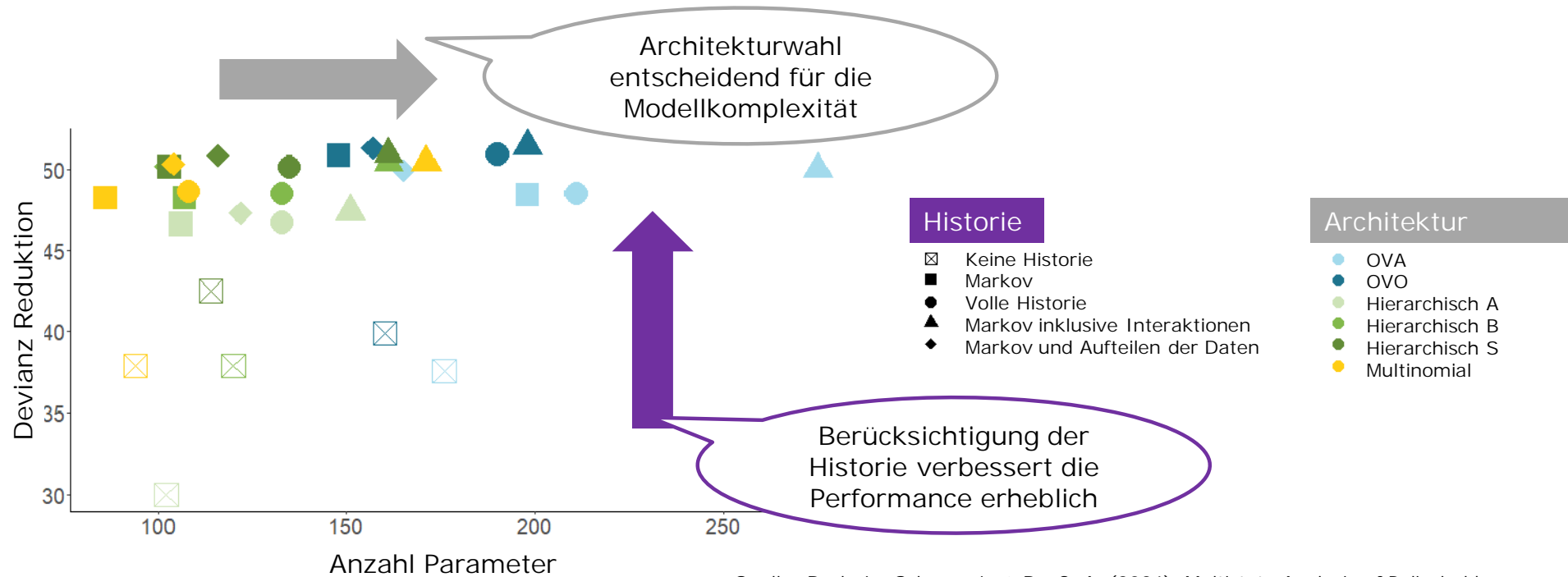
Mehrere VN-Optionen - Designentscheidungen

- weitere Designentscheidungen zu treffen (für m Ausprägungen der Zielgröße)
- **Qualitativ:**
 - Varianten der Modellarchitektur
 - Zerlegung in mehrere **binäre Lasso-Modelle**
 - OVA → einfach (m eindeutige Modelle)
 - OVO → komplexe Aggregation ($\frac{m(m-1)}{2}$ eindeutige Modelle)
 - Hierarchisch → einfach (m-1 Modelle aber $\frac{2^{m-1}\Gamma(m-\frac{1}{2})}{\sqrt{\pi}}$ - Varianten wie diese aussehen)
 - **Multinomiales Lasso-Modell**
 - Implementierungen sind hier nur noch begrenzt vorhanden
 - Glättung über Merkmale hinweg geht; Glättung über Ausprägungen der Zielgröße geht bisher nicht
 - Berücksichtigung der Historie
 - keine Historie / aktueller Zustand / volle Historie



Weitere Konfigurationsmöglichkeiten und Anwendungen

Mehrere VN-Optionen – quantitative Ergebnisse



Quelle: Reck, L., Schupp, J., & Reuß, A. (2024). Multistate Analysis of Policyholder Behaviour in Life Insurance-Lasso based Modelling Approaches. Working Paper, Ulm University.

Fazit

Das Lasso zur Automatisierung aktuarieller Analysen



multivariater, performanter, automatisierter Ansatz
vielfältige, interpretierbare Strukturen innerhalb der Kovariablen identifizierbar
große Bandbreite an Verteilungen der Zielgröße möglich (z.B. multinomial)



Designentscheidungen („Customizing“) heben letztlich das Potenzial des Modells.
Wie bei jedem Modell ist die Kenntnis der Stärken UND Schwächen des Modells wichtig!



Monitoring biometrischer Annahmen (Tod, Invalidisierung, Reaktivierung, ...) → Identifikation auffälliger Segmente
Prognosemodelle im Leistungsbearbeitungsprozess (z.B. Anerkennungswahrscheinlichkeiten in der BU)
VN-Optionen in anderen Sparten (z.B. Stornotreppen in KFZ)



Bei vergleichbaren Ausgangslagen ist Lasso immer eine Methode, die man berücksichtigen sollte, d.h.
- bei mehreren Kovariablen, mit unterschiedlichen (oder unbekannt) Strukturen,
- wenn die Interpretierbarkeit des Modells wichtig ist.

Kontakt



Dr. Johannes Schupp
j.schupp@ifa-ulm.de



Dr. Lucas Reck
l.reck@ifa-ulm.de

