

Tarifierung in der Schaden-/Unfallversicherung

Ergänzung klassischer Tarifierung durch moderne Data-Analytics-Methoden

Lukas Hahn
Institut für Finanz- und Aktuarwissenschaften (ifa), Ulm



DAV

DEUTSCHE
AKTUARVEREINIGUNG e.V.

Institut für Finanz- und
Aktuarwissenschaften

ifa



DAV-Jahrestagung 2019
Düsseldorf, 25.04.2019

Ergänzung klassischer Tarifierung durch moderne Data-Analytics-Methoden

Agenda

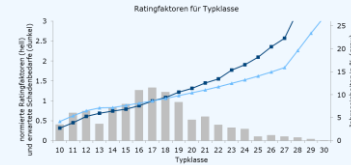
Einleitung	2
Alternative zur Black Box: GLM-Regularisierung	6
Fallstudie Kfz-Haftpflicht	12
Wrapup	17

In a Nutshell

Klassischer Tarifierungsprozess am Beispiel K/KH



Nettorisikoprämie
= Basisprämie * Ratingfaktoren



Risikomodellierung

iterative Anpassung multivariater Modelle zur bestmöglichen Schätzung einzelvertraglicher Schadenerwartungen auf dem Modellbestand

typisch:
GLM

Tarifmodellierung und Übergang zum Tarifbuch

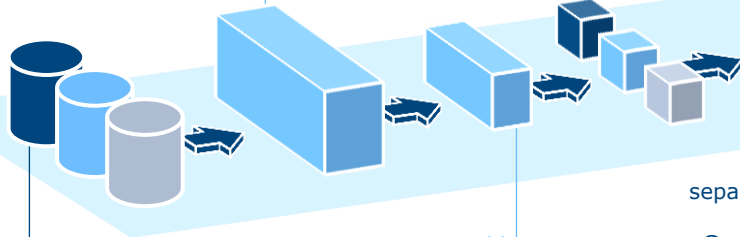
Tarifbestand, Kosten und Marge, Marktpositionierung
rechtlicher und regulatorischer Rahmen
Tariffaktoren: Spreizung, Dämpfung, Fixierung, Organik, ...



Datenzusammenführung:
Analyse, Validierung,
Bereinigung, Anpassung, ...



Bausteine:
Sparten
Schadenarten
Schadenkennzahlen



Datenquellen

Vorauswahl
der Merkmale

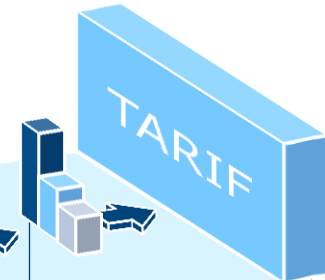
separate Modellierung:
z. B. Trends und
Großschadenkappung



Niveaueinstellung
Bestandsveränderung
Leistungsveränderung
Kosten
Puffer und Marge
Rabatte



Benchmarking:
Marktvergleiche
Tarifbewertung



Erstellung
Tarifbuch

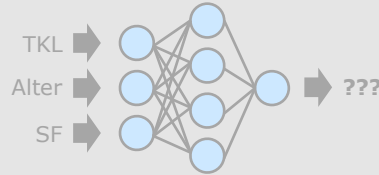




Black-Box-Ansatz?

... mit Risiken und Nebenwirkungen

Nettorisikoprämie
= Output der Black Box



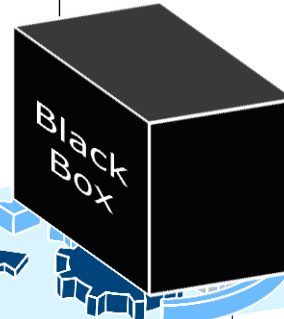
Risikomodellierung
durch komplexen
Algorithmus



Datenzusammenführung:
Analyse, Validierung,
Bereinigung, Anpassung, ...



Bausteine:
Sparten
Schadenarten
Schadenkennzahlen



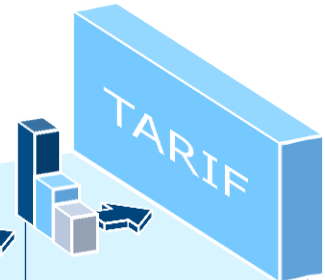
separate Modellierung:
z. B. Trends und
Großschadenkappung



Tarifmodellierung und Übergang zum Tarifbuch?
Spreizung und Dämpfung einzelner Tarifmerkmale?
Sicherstellung der Organik?
Erfüllung rechtlicher und regulatorische Anforderungen?
→ Akzeptanz?



Niveaunpassung
Bestandsveränderung
Leistungsveränderung
Kosten
Puffer und Margen
Rabatte



Erstellung
Tarifbuch



Benchmarking:
Marktvergleiche
Tarifbewertung



Datenquellen



Vorauswahl
der Merkmale



Dann White Box du nutzen musst!

Alternative mittels GLM-Regularisierung

Ziel **simultane** Anpassung aller Merkmale



datenbasiertes Erlernen von Strukturen

automatisierter Prozess

nachvollziehbares Modell

nahtlose Einbindung in Prozesse



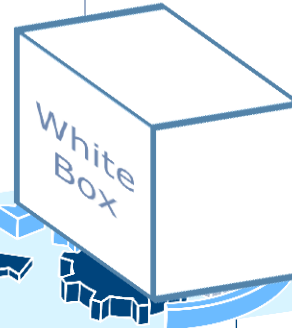
Risikomodellierung
durch Regularisierung
eines GLM



Datenzusammenführung:
Analyse, Validierung,
Bereinigung, Anpassung, ...



Bausteine:
Sparten
Schadenarten
Schadenkennzahlen

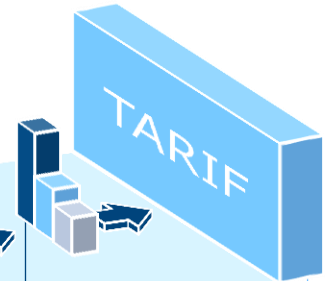


separate Modellierung:
z. B. Trends und
Großschadenkappung



Tarifmodellierung und Übergang zum Tariffbuch

Tarifbestand, Kosten und Margen, Marktpositionierung
rechtlicher und regulatorischer Rahmen
Tariffaktoren: Spreizung, Dämpfung, Fixierung, Organik, ...



Erstellung
Tariffbuch



Benchmarking:
Marktvergleiche
Tariffbewertung



Ergänzung klassischer Tarifierung durch moderne Data-Analytics-Methoden

Agenda

Einleitung	2
Alternative zur Black Box: GLM-Regularisierung	6
Fallstudie Kfz-Haftpflicht	12
Wrapup	17

Ausgangslage

GLMs optimieren eine Likelihoodfunktion

Definition: Generalized Linear Model (GLM)



Verteilungsannahme

$$f(Y; \theta, \phi) = \exp \left(\frac{Y\theta - b(\theta)}{\phi} + c(Y; \phi) \right)$$

Linkfunktion

$$g: \mathbb{R} \rightarrow \mathbb{R} \text{ sodass } \eta = g(\mu) \text{ mit } \mu = \mathbb{E}(Y)$$

linearer Prädiktor

$$\eta = X\beta = \beta_0 + \beta_1 X_1 + \dots + \beta_m X_m$$

- Modellanpassung:

- Schätzung von β mit Maximum-Likelihood-Methode

$$\ell(\beta; y_i, x_i) = -\frac{1}{n} \sum_{i=1}^n \log f(y_i \mid x_i, \beta) \rightarrow \min$$

- iterativ mittels Fisher-Scoring-Algorithmus

Regularisierung im GLM

Bestrafung der Komplexität

- Bestrafungsterm für zu hohe Komplexität
 - Komplexität $\approx$ zu viele und/oder zu große Parameter
 - Vermeidung von Overfitting

$$\ell_{\text{pen}}(\beta; y_i, x_i, \lambda) = \ell(\beta; y_i, x_i) + \lambda \Omega(\beta_1, \dots, \beta_m) \rightarrow \min$$

- Komponenten:

Shrinkage-Faktor $\lambda \geq 0$ zum Ausgleich zwischen

- Bestrafung $\Omega(\beta_1, \dots, \beta_m)$ der Komplexität und
- Likelihood-Anpassung $\ell(\beta; y_i, x_i)$ an die Daten

Bestrafungsfunktion $\Omega: \mathbb{R}^m \rightarrow \mathbb{R}_+$
für die Parameter $\beta_1, \dots, \beta_m$
(ohne Intercept)

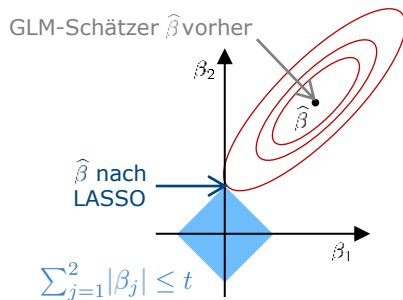
technische Voraussetzung:

Standardisierung der Merkmale x_i zur Vergleichbarkeit der β notwendig

LASSO und Ridge Mischen possible

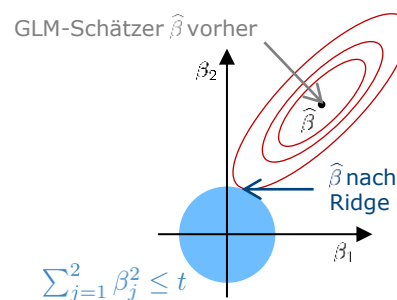
LASSO (Tibshirani 1996)¹ Bestrafung der ℓ_1 -Norm

$$\Omega(\beta_1, \dots, \beta_m) = \|(\beta_1, \dots, \beta_m)\|_1 = \sum_{j=1}^m |\beta_j|$$



Ridge (Hoerl 1970)² Bestrafung der ℓ_2 -Norm

$$\Omega(\beta_1, \dots, \beta_m) = \|(\beta_1, \dots, \beta_m)\|_2^2 = \sum_{j=1}^m \beta_j^2$$



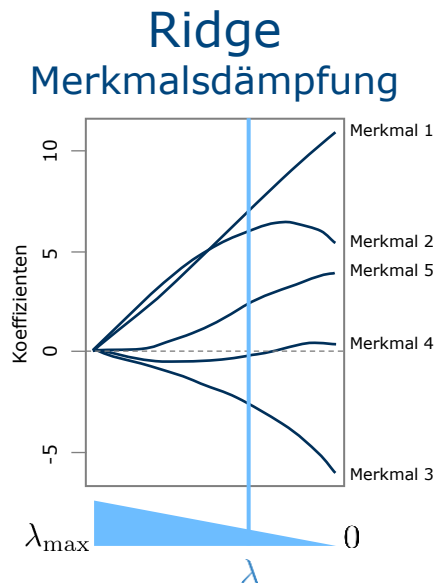
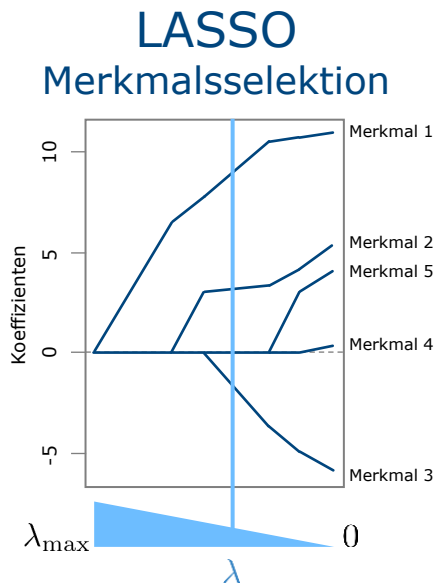
Elastisches Netz (Zou/Hastie 2005): Mischung beider Normen

$$\Omega(\beta_1, \dots, \beta_m; \alpha) = \alpha \|(\beta_1, \dots, \beta_m)\|_1 + \frac{1}{2}(1 - \alpha) \|(\beta_1, \dots, \beta_m)\|_2^2$$

¹ Least Absolute Selection and Shrinkage Operator
² engl. „Bergrücken“; benannt in Anlehnung zu Ridge-Analysen (Hoerl 1970), auch Thikonov-Regularisierung

Eigenschaften

Lambda-Pfade



Elastische Netze kombinieren
beide Eigenschaften!

Verschiedene Merkmalstypen...

... verschiedene Varianten

	Standard LASSO	Group LASSO	Fused LASSO
Bestrafungsfunktion	$\sum_{j=1}^m \beta_j $	$\sum_{j=1}^m \sqrt{\sum_{k=1}^{m_j} \beta_{j,k}^2}$	$\sum_{j=1}^m \sum_{k=2}^{m_j} \beta_{j,k} - \beta_{j,k-1} $
Lambda-Pfad			
Varianten	Elastic Net Relaxed LASSO Weighted LASSO	Overlap Group LASSO Sparse Group LASSO	2D-Fused LASSO Generalized Fused LASSO Trend Filtering
Eignung	Merkmalstypen Modellierung	stetig binär Haupteffekte	nominal ordinal spatial Gruppieren Hierarchien Binning Interaktionen

Multi-type LASSO (mit Bestrafungsbudgets)

$$\ell_{\text{pen}}(\beta; y_i, x_i, \lambda) = \ell(\beta; y_i, x_i) + \sum_{j=1}^m \lambda_j \Omega_j(\beta_j)$$

Ergänzung klassischer Tarifierung durch moderne Data-Analytics-Methoden

Agenda

Einleitung	2
Alternative zur Black Box: GLM-Regularisierung	6
Fallstudie Kfz-Haftpflicht	12
Wrapup	17

Ein Ansatz sie alle zu einem zum Umgang mit allen Arten von Merkmalen



Idee: Regularisierung einer extrem großen Design-Matrix

- geschickte Dummy-Codierung zum adaptiven Erlernen nicht-linearer Strukturen
- Haupteffekte und Interaktionen berücksichtigt
- Zu komplex?
→ Lass das mal das LASSO machen!



Alter (VN)	...	41	42	43	44	45	...
43-Jähriger	...	1	1	1	0	0	...
44-Jähriger	...	1	1	1	1	0	...

Beispiel: „Kontrast-Codierung“ für eine Plateau-Modellierung



Optionen: verbleibende Modelleinstellungen

- übergreifendes Strafbudget λ
- Mischungsparameter α für elastisches Netz (Kontrolle der Multikollinearität)
- Auswahl für Anpassung (pro Merkmal): Plateaus, Linearitäten oder Hybrid

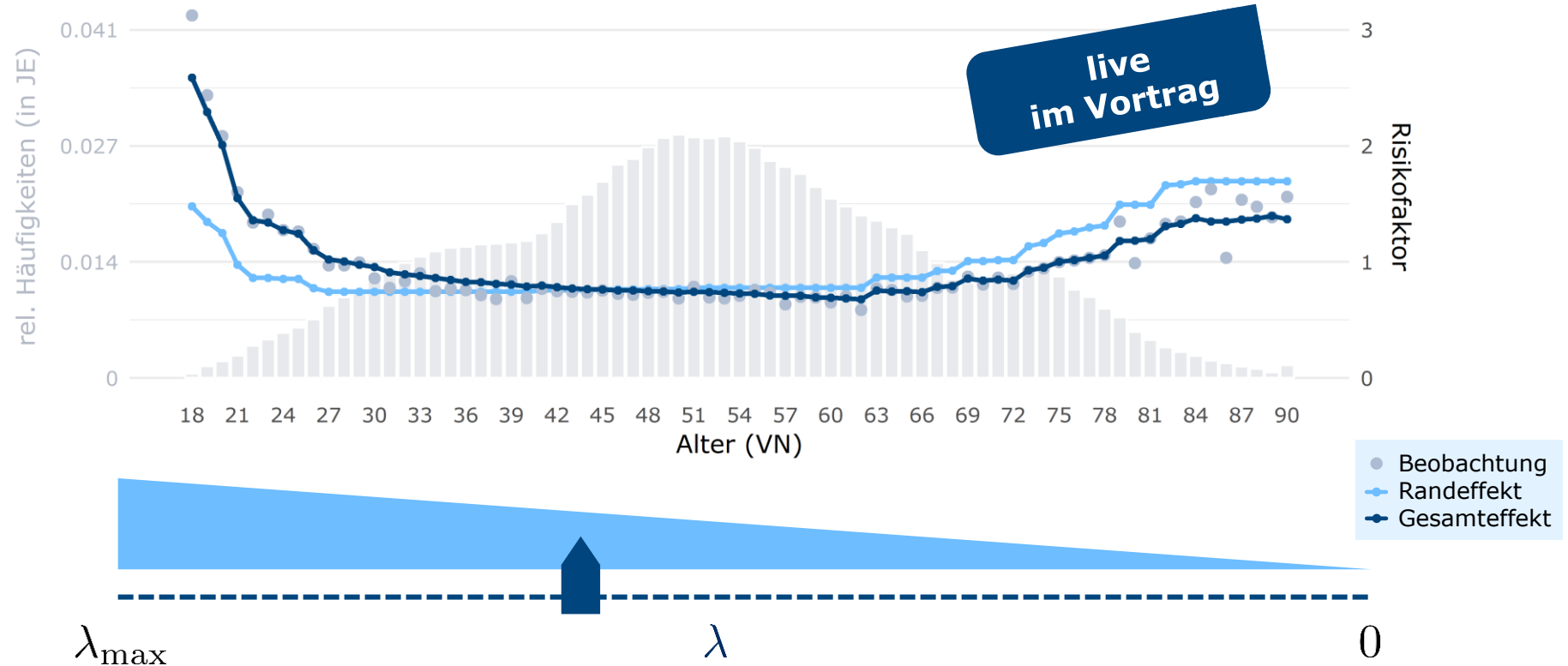


Ergebnis: datengetriebene GLM-Modellierung

- simultane und einheitliche Berücksichtigung aller Merkmale unterschiedlicher Art
- objektiv datengetriebener und verstärkt automatisierter Modellierungsprozess
- integrierte Merkmalsselektion, Binning, Transformationen, Interaktionen

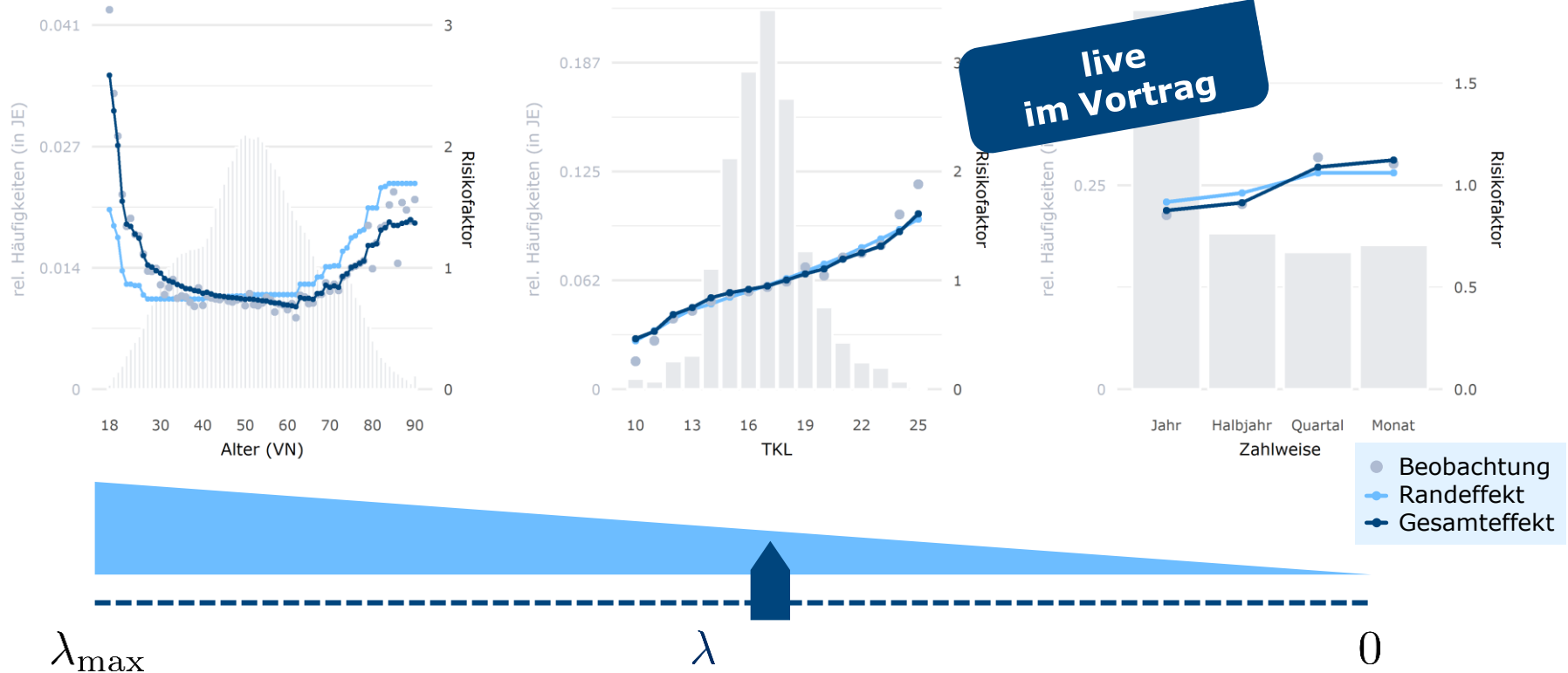
DAV-Jahrestagung 2019, Düsseldorf, 25.04.2019

Beispiel: adaptive Anpassung für das Alter (VN) unter simultaner Anpassung der Ausgleichseffekte aller anderen Merkmale



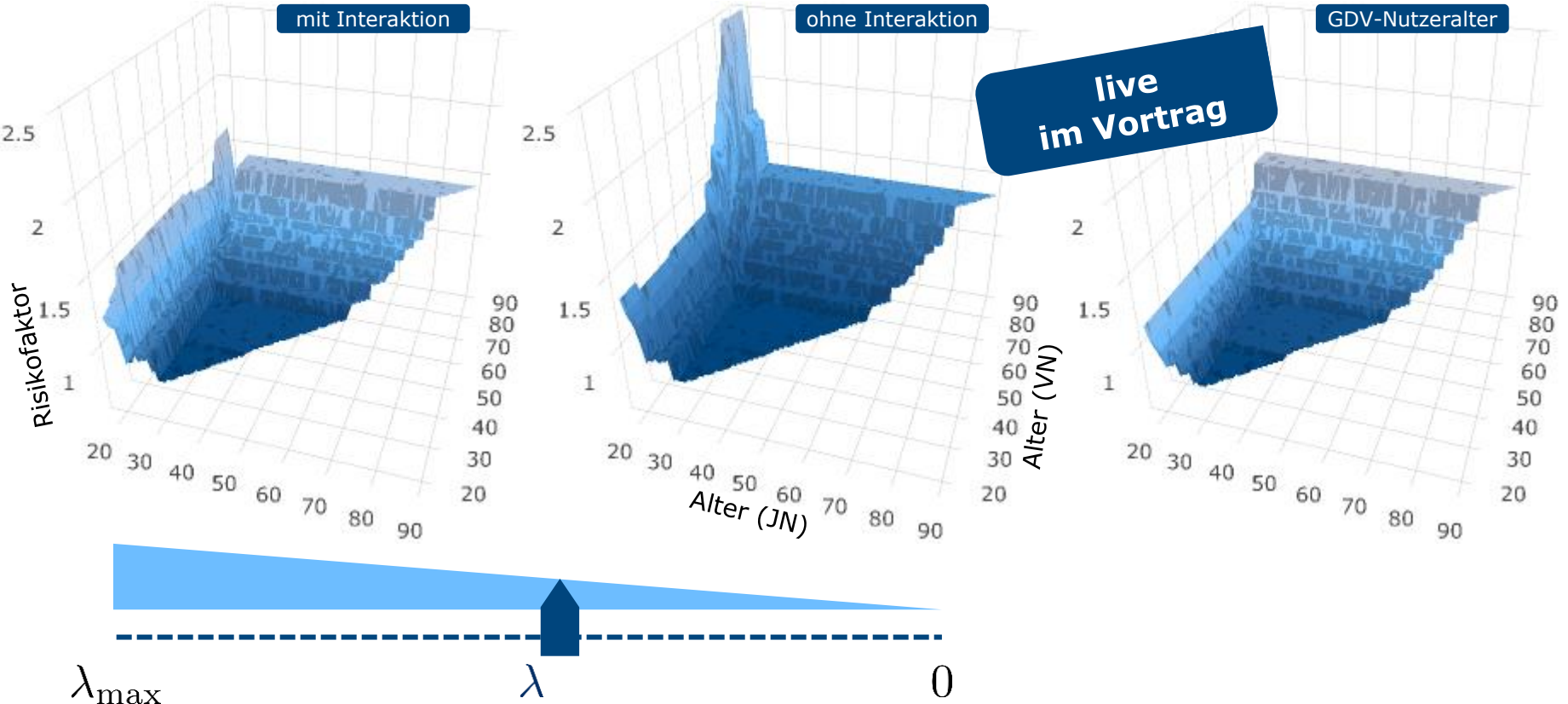
DAV-Jahrestagung 2019, Düsseldorf, 25.04.2019

Beispiel: gleichzeitige Anpassung stetiger und ordinaler Merkmale unter simultaner Anpassung der Ausgleichseffekte aller anderen Merkmale



DAV-Jahrestagung 2019, Düsseldorf, 25.04.2019

Beispiel: adaptive Anpassung eines Interaktionseffekts unter simultaner Anpassung der Ausgleichseffekte aller anderen Merkmale



Ergänzung klassischer Tarifierung durch moderne Data-Analytics-Methoden

Agenda

Einleitung	2
Alternative zur Black Box: GLM-Regularisierung	6
Fallstudie Kfz-Haftpflicht	12
Wrapup	17

Wrapup

17 Folien auf den Punkt gebracht

Sie sahen: Risikomodellierung mit elastischen Netzen unter Berücksichtigung aller Arten von Merkmalen



Vorauswahl und Vorverarbeitung der Merkmale

manuelle und iterative Modellierung

sukzessive und individuelle Modellierung pro Merkmal

integrierte Merkmalsselektion und -vorverarbeitung

simultane, datengetriebene und automatisierte Merkmalsanpassung

verschiedene Merkmalstypen, implizite Wirkung der Ausgleichseffekte sowie adaptive Erkennung nicht-linearer Strukturen

nachvollziehbarer Modelloutput in bewährter GLM-Struktur

gezielte Kontrolle über zentrale Freiheitsgrade beim Anwender

globales Strafbudget

Mischungsparameter

Plateau vs. Linearität

nahtlose Integration in bestehende Tarifierungsprozesse

für alle Sparten geeignet – und ganz im Gegenteil: **effiziente GLM-Erstellung gerade dort, wo die Risikomodelle noch nicht so komplex sind!**



absolute Black Box

komplexes Tuning notwendig

Einsatz in der Tarifmodellierung noch unklar



mal wieder reine K/KH-Anwendung



Institut für Finanz- und Aktuarwissenschaften

Kontaktdaten und Beratungsangebot

Life



Produktentwicklung
Biometrische Risiken
Zweitmarkt

Non-Life



Produktentwicklung und
Tarifizierung
Schadenreservierung
Risikomodellierung

Health



Aktuarieller
Unternehmenszins
Leistungsmanagement

Actuarial Consulting

Solvency II • Embedded Value • Asset-Liability-Management
ERM • wert- und risikoorientierte Steuerung • Data Analytics

Projektmanagement • Markteintritt • Bestandsmanagement • strategische Beratung

Actuarial Services

aktuarielle Großprojekte • aktuarielle Tests
Überbrückung von Kapazitätsengpässen

Research



Aus- und Weiterbildung



... weitere Informationen unter www.ifa-ulm.de

Lukas Hahn
+49 731 20644-239
l.hahn@ifa-ulm.de



Quellen

Literatur und Bildnachweise

• Literatur

- Devriendt, S., Antonio, K., Reynkens, T., & Verbelen, R. (2018). Sparse Regression with Multi-type Regularized Feature Modeling. arXiv preprint arXiv:1810.03136.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: prediction, inference and data mining. Springer-Verlag, New York.
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.
- Tibshirani, R., Wainwright, M., & Hastie, T. (2015). Statistical learning with sparsity: the lasso and generalizations. Chapman and Hall/CRC.
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2), 301-320.

• Bildnachweise

- Folien 3-5, 11, 13-16, 18: eigene Darstellung; all icons made by Freepik from www.flaticon.com.
- Folie 7: eigene Darstellung; designed by Showeet.com.
- Folien 9-10: eigene Darstellung in Anlehnung an Tibshirani, R., Wainwright, M., & Hastie, T. (2015). Statistical learning with sparsity: the lasso and generalizations. Chapman and Hall/CRC.