A Pricing App Using Regularized GLMs to Enhance Risk Modeling in P&C

Dr. Lukas Hahn
SV SparkassenVersicherung, Stuttgart, Germany

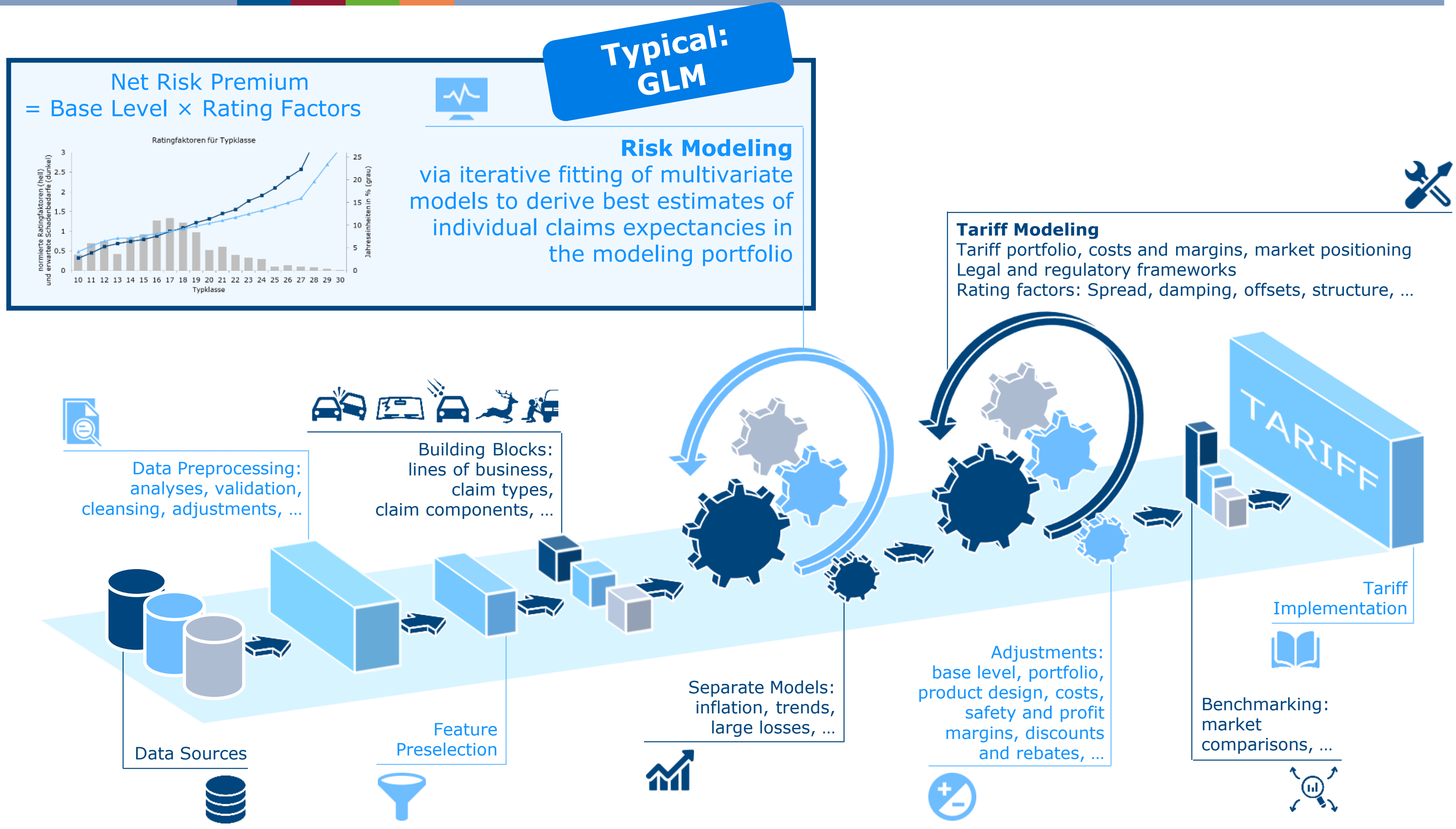
DGVFM
DEUTSCHE GESELLSCHAFT FÜR VERSICHERUNGS- UND FINANZMATHEMATIK e.V.

eWeiterbildungstag der DGVFM, 29 October 2020

eWeiterbildungstag der DGVFM, 29 October 2020

DGVFM
DEUTSCHE GESELLSCHAFT FÜR VERSICHERUNGS- UND FINANZMATHEMATIK e.V.

In a Nutshell
A Classical Pricing Process in Motor



Source: Lukas Hahn, "Tarifierung in der Schaden-/Unfallversicherung - Ergänzung klassischer Tarifierung durch moderne Data-Analytics-Methoden", DAV Annual Conference 2019 in Düsseldorf, Germany, on 25 April 2019, Section "Actuarial Data Science", https://aktuar.de/veranstaltungen/unsere-veranstaltungen/uebersicht/_layouts/15/MyAccount/CongressDetails.aspx?eventno=REC19JTG

DGVFM
DEUTSCHE GESELLSCHAFT FÜR VERSICHERUNGS- UND FINANZMATHEMATIK e.V.

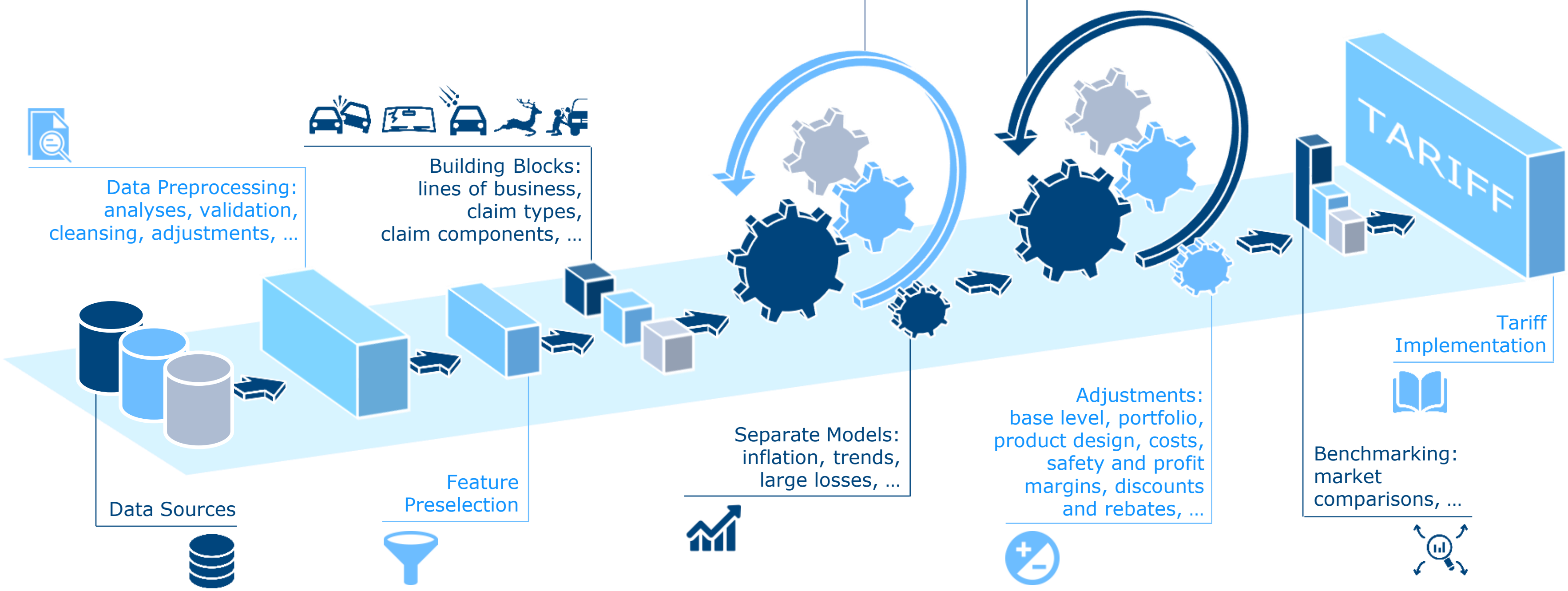
In a Nutshell
A Classical Pricing Process in Motor

Target

- **Simultaneous** fit of all risk variables
- **Data-driven** adaptive learning of patterns
- **Automized** process
- **Explainable** model
- **Seamless** integration in existing processes

Risk Modeling
via regularized GLMs

Tariff Modeling
Tariff portfolio, costs and margins, market positioning
Legal and regulatory frameworks
Rating factors: Spread, damping, offsets, structure, …



Source: Lukas Hahn, "Tarifierung in der Schaden-/Unfallversicherung - Ergänzung klassischer Tarifierung durch moderne Data-Analytics-Methoden", DAV Annual Conference 2019 in Düsseldorf, Germany, on 25 April 2019, Section "Actuarial Data Science", https://aktuar.de/veranstaltungen/unsere-veranstaltungen/uebersicht/_layouts/15/MyAccount/CongressDetails.aspx?eventno=REC19JTG

GLM Regularization

Punish the Complexity!

Definition: Generalized Linear Model (GLM)

Probability distribution	$f(Y;\theta,\phi) = \exp\left(\frac{Y\theta - b(\theta)}{\phi} + c(Y;\phi)\right)$
Link function	$g\colon \mathbb{R} \to \mathbb{R}$ s.th. $\eta = g(\mu)$ with $\mu = \mathbb{E}(Y)$
Linear predictor	$\eta = X\beta = \beta_0 + \beta_1 X_1 + \ldots + \beta_m X_m$

Standard

Estimate β using maximum likelihood:

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

Extra

Introduce penalty term to punish complexity (reduce overfitting):

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

Components:

Regularization Parameter $\lambda \geq 0$ to balance
- punishment $\Omega(\beta_1, \ldots, \beta_m)$ of complexity and
- fit of likelihood $\ell(\beta; y_i, x_i)$ to the data

Penalty Function $\Omega\colon \mathbb{R}^m \to \mathbb{R}_+$
for parameters $\beta_1, \ldots, \beta_m$
(without intercept)

Technical prerequisites: Standardize variables x_i to allow for comparison of the magnitude of β

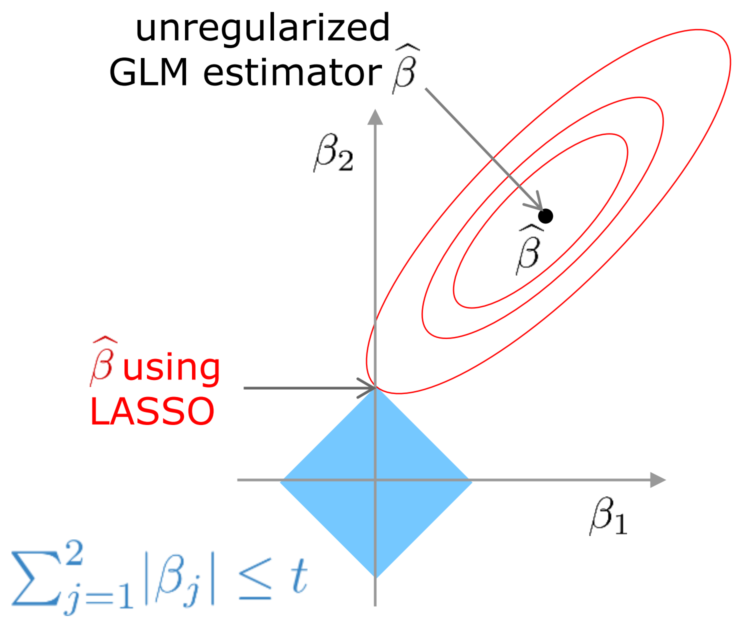
Source: Lukas Hahn, "Tarifierung in der Schaden-/Unfallversicherung - Ergänzung klassischer Tarifierung durch moderne Data-Analytics-Methoden", DAV Annual Conference 2019 in Düsseldorf, Germany, on 25 April 2019, Section "Actuarial Data Science", https://aktuar.de/veranstaltungen/unsere-veranstaltungen/uebersicht/_layouts/15/MyAccount/CongressDetails.aspx?eventno=REC19JTG

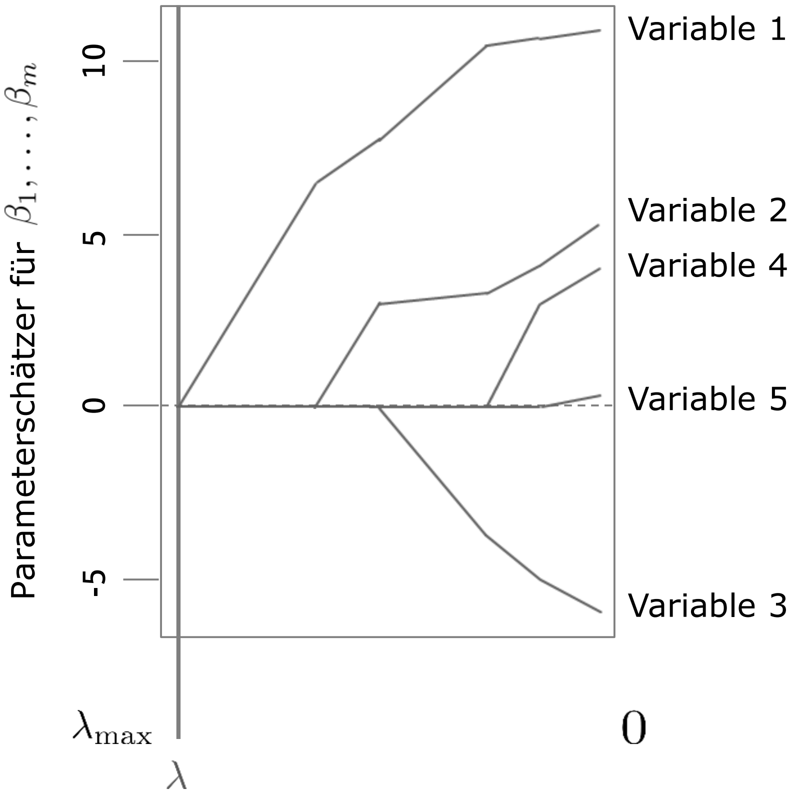
Now it is Time for an Example
The LASSO

LASSO: **L**east **A**bsolute **S**election and **S**hrinkage **O**perator (Tibshirani, 1996)

- Penalty Function: ℓ_1 norm

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





Source: Lukas Hahn, "Tarifierung in der Schaden-/Unfallversicherung - Ergänzung klassischer Tarifierung durch moderne Data-Analytics-Methoden", DAV Annual Conference 2019 in Düsseldorf, Germany, on 25 April 2019, Section "Actuarial Data Science", https://aktuar.de/veranstaltungen/unsere-veranstaltungen/uebersicht/_layouts/15/MyAccount/CongressDetails.aspx?eventno=REC19JTG

One Method to Rule Them All

Allowing for All Types of Variables

- The pricing app makes use of a regularization technique that allows for all types of variables.
- The method was first proposed by the Institute for Financial and Actuarial Sciences (ifa) at the DAV Annual Conference 2019 in Düsseldorf, Germany.

Main idea (ifa): Regularization of an extremely large design matrix

- Use clever dummy/contrast coding to enable adaptive learning of non-linear structures
- Allow for main effects and interactions
- You think this is too complex?
  ➔ Bring in the LASSO!

Age (PH)	…	41	42	43	44	45	…
43-year-old	…	1	1	1	0	0	…
44-year-old	…	1	1	1	1	0	…

Example: Contrast coding for staircase functions

Options: Remaining model calibration

- Global penalty budget λ
- Coding setup per variable: staircase, stepwise linear, or hybrid
- Optional: Manual adjustments per variable

Result: Data-driven GLM modeling

- Simultaneous and coherent allowance for variables of different types
- Objectively data-driven and increasingly automized modeling process
- Integrated feature selection, binning, transformations, interactions, …
- Well-known model output for visualization, interpretation, and integration in existing pricing processes

Source: Lukas Hahn, "Tarifierung in der Schaden-/Unfallversicherung - Ergänzung klassischer Tarifierung durch moderne Data-Analytics-Methoden", DAV Annual Conference 2019 in Düsseldorf, Germany, on 25 April 2019, Section "Actuarial Data Science", https://aktuar.de/veranstaltungen/unsere-veranstaltungen/uebersicht/_layouts/15/MyAccount/CongressDetails.aspx?eventno=REC19JTG

Live Demo
Motor Case Study with Realistic Data

Live Demo

Sparkassen Versicherung

Thank you for your attention!

Dr. Lukas Hahn

SV SparkassenVersicherung Stuttgart
Löwentorstraße 65
70376 Stuttgart

lukas.hahn@sparkassenversicherung.de
+49 711 898-43765

References
Literature and Credits

- Literature

  - 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.

- Credits
  - All illustrations by courtesy of the Institute for Financial and Actuarial Sciences (ifa) in Ulm, Germany; taken from Lukas Hahn, "Tarifierung in der Schaden-/Unfallversicherung - Ergänzung klassischer Tarifierung durch moderne Data-Analytics-Methoden", DAV Annual Conference 2019 in Düsseldorf, Germany, on 25 April 2019, Section "Actuarial Data Science"
  - In addition,
    - for slides 2 and 5: All icons made by Freepik from www.flaticon.com.
    - for slide 3: All icons designed by Showeet.com.
    - for slide 4: Own illustration based on Tibshirani, R., Wainwright, M., & Hastie, T. (2015). Statistical learning with sparsity: the lasso and generalizations. Chapman and Hall/CRC.