Auto-Calibration and Isotonic Recalibration

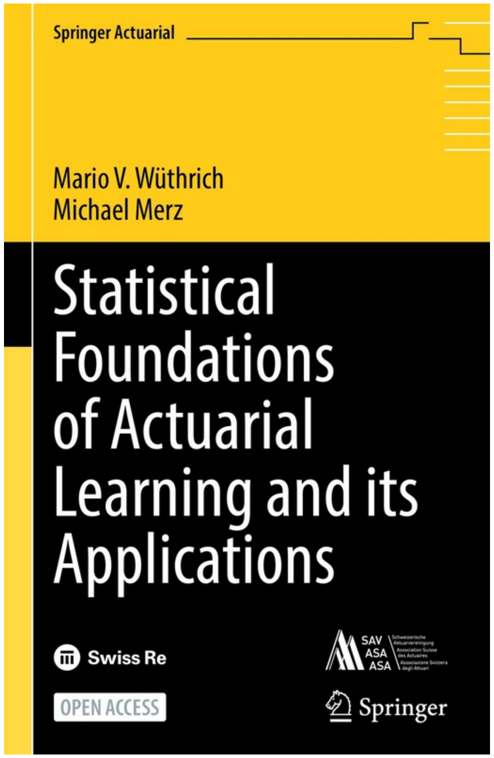
Mario V. Wüthrich
RiskLab, ETH Zurich

Joint work with Johanna Ziegel

May 24, 2023
Actuarial Data Science Après-Midi
Schweizerische Aktuarvereinigung SAV, Zurich

Overview

- Introduction: regression models
- Auto-calibration and isotonic recalibration
- Score decomposition
- Conclusions



- **Introduction: regression models**

Best-estimate/pure risk premium

- Assume we have independent data points $(Y_i, \boldsymbol{x}_i)_{i=1}^n$ that have been generated by a fixed but unknown model:

  ⋆ Y_i are the responses (insurance claims), and
  ⋆ $\boldsymbol{x}_i$ are the covariates (features, explanatory variables, independent variables).

- For known data generating model, one can compute the (conditionally) expected claim of each policyholder with covariates $\boldsymbol{x}$ by

$$\mu^*(\boldsymbol{x}) := \mathbb{E}\left[Y \middle| \boldsymbol{x}\right].$$

- $\mu^*(\boldsymbol{x})$ is the true best-estimate/pure risk premium for claim Y of policyholder $\boldsymbol{x}$.

- For unknown data generating model, one needs to estimate μ^* from the observed sample $(Y_i, \boldsymbol{x}_i)_{i=1}^n$.

Regression model approach

- **Problem:** Estimate from sample $(Y_i, \boldsymbol{x}_i)_{i=1}^n$ the true regression function

$$\boldsymbol{x} \mapsto \mu^*(\boldsymbol{x}) = \mathbb{E}\left[Y \mid \boldsymbol{x}\right].$$

(1) Select a class of regression functions (models)

$$\mathcal{M} = \Big\{ \boldsymbol{x} \mapsto \mu(\boldsymbol{x});\ \mu \text{ has certain properties} \Big\},$$

as candidates to approximate μ^* (which can be any $\sigma(\boldsymbol{x})$-measurable function).

(2) Solve the minimization problem

$$\widehat{\mu} = \underset{\mu \in \mathcal{M}}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} L\left(Y_i, \mu(\boldsymbol{x}_i)\right),$$

for a strictly consistent loss function L for mean estimation; see Gneiting (2011).

Quality of approximation

- **Question:** Is the selected model $\widehat{\mu} \in \mathcal{M}$ a good approximation to μ^*?

- Generally, this depends on: the properties of μ^*, the selected class $\mathcal{M}$ of candidates, the sample $(Y_i, \boldsymbol{x}_i)_{i=1}^n$, the loss function L and the optimization algorithm used.

- Because the true regression model μ^* is unknown, this question is analyzed empirically using out-of-sample data, providing an (independent) empirical approximation to the true data generating model.

- Popular quality and goodness-of-fit criteria:

  ⋆ out-of-sample losses (under strictly consistent loss functions);
  ⋆ Gini indices (under auto-calibration);
  ⋆ auto-calibration and the balance property;
  ⋆ conditional T-reliability diagrams and score decomposition.

- **Auto-calibration and isotonic recalibration**

Auto-calibration

- In a binary context, auto-calibration was used in Hosmer–Lemeshow (1980), Schervish (1989), Menon et al. (2012), Tasche (2021), and, more generally, in Tsyplakov (2013), Gneiting–Ranjan (2013), Pohle (2020), Krüger–Ziegel (2021), Denuit et al. (2021), Fissler et al. (2022), Gneiting–Resin (2022), W. (2023), Lindholm et al. (2023), W.–Ziegel (2023).

The regression function $\boldsymbol{x} \mapsto \mu(\boldsymbol{x})$ is auto-calibrated for $(Y, \boldsymbol{x})$ if, a.s.,

$$\mu(\boldsymbol{x}) = \mathbb{E}\left[Y \,|\, \mu(\boldsymbol{x})\right].$$

▷ Intuitively, this means that every price cohort $\mu(\boldsymbol{x})$ is in average self-financing,

or in other words, there is no systematic cross-financing between different price cohorts $\mu(\boldsymbol{x}) \neq \mu(\boldsymbol{x}')$ within the portfolio.

▷ Insurance price systems should generally fulfill this important property!

Empirical testing for auto-calibration

The regression function $\boldsymbol{x} \mapsto \mu(\boldsymbol{x})$ is auto-calibrated for $(Y, \boldsymbol{x})$ if, a.s.,

$$\mu(\boldsymbol{x}) = \mathbb{E}\left[Y \middle| \mu(\boldsymbol{x})\right].$$

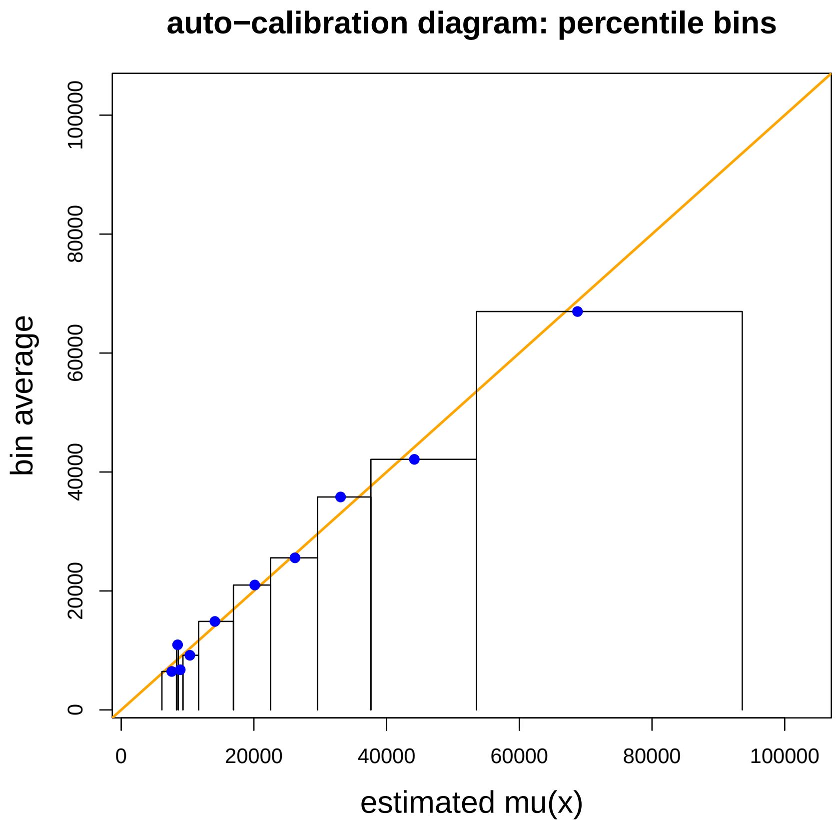
- **Binning** (Hosmer–Lemeshow (1980) χ^2-test): Build disjoint price intervals (bins) $I_k = [a_k, a_{k+1})$ and consider (out-of-sample) the average claim in each bin

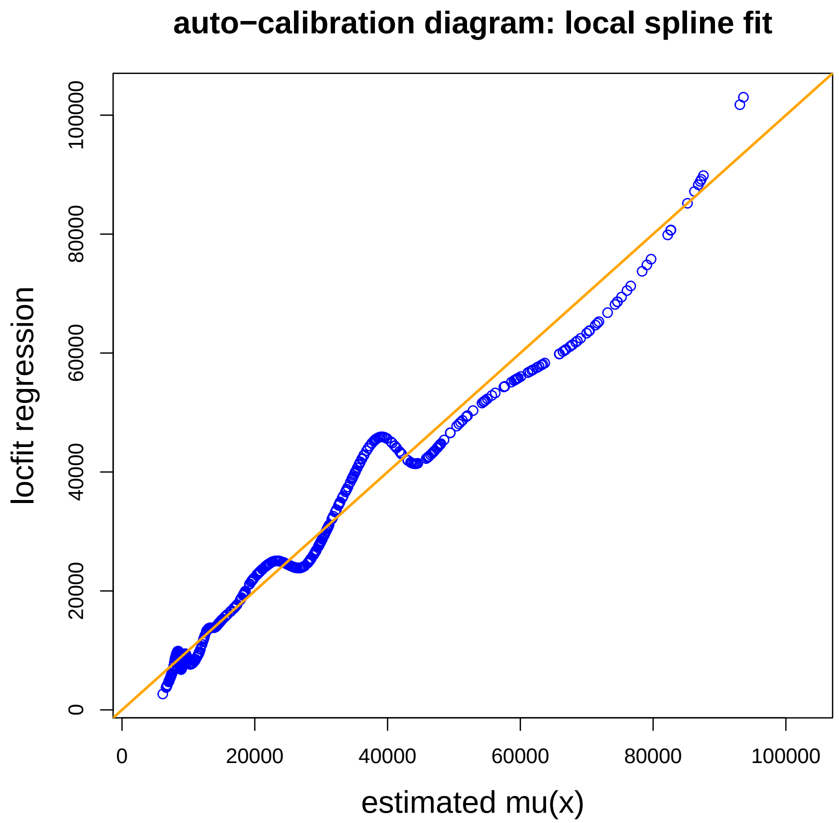
$$\frac{\sum_{i=1}^n \mu(\boldsymbol{x}_i)\, \mathbb{1}_{\{\mu(\boldsymbol{x}_i) \in I_k\}}}{\sum_{i=1}^n \mathbb{1}_{\{\mu(\boldsymbol{x}_i) \in I_k\}}} \overset{??}{\approx} \frac{\sum_{i=1}^n Y_i\, \mathbb{1}_{\{\mu(\boldsymbol{x}_i) \in I_k\}}}{\sum_{i=1}^n \mathbb{1}_{\{\mu(\boldsymbol{x}_i) \in I_k\}}}.$$

- **Local Regression** (Loader (1999)): Binning is a discretized version of a local regression (smoother) that regresses the responses Y from the price cohorts $\mu(\boldsymbol{x})$

$$\texttt{locfit}(Y \sim \mu(\boldsymbol{x}),\ \texttt{alpha} = 0.1,\ \texttt{deg} = 2).$$

Lift plots for empirical auto-calibration tests





(left): decile bins (right): local spline regression

Isotonic recalibration (1/2)

- Isotonic regression is a non-parametric way to restore the auto-calibration property; a mathematical argument follows in Eq. (1), below.

Isotonic regression solves the following optimization problem (for positive weights w_i)

$$\widehat{\boldsymbol{m}} = \underset{\boldsymbol{m}=(m_1,\ldots,m_n)^\top \in \mathbb{R}^n}{\arg\min} \sum_{i=1}^{n} w_i \left(Y_i - m_i\right)^2,$$

subject to $m_k \leq m_j \iff \mu(\boldsymbol{x}_k) \leq \mu(\boldsymbol{x}_j)$.

- Isotonic regression preserves the ordering in $\mu(\boldsymbol{x}_i)_{i=1}^n$. This requires that the first regression function $\mu(\cdot)$ provides (approximately) the correct ordering.

Isotonic recalibration (2/2)

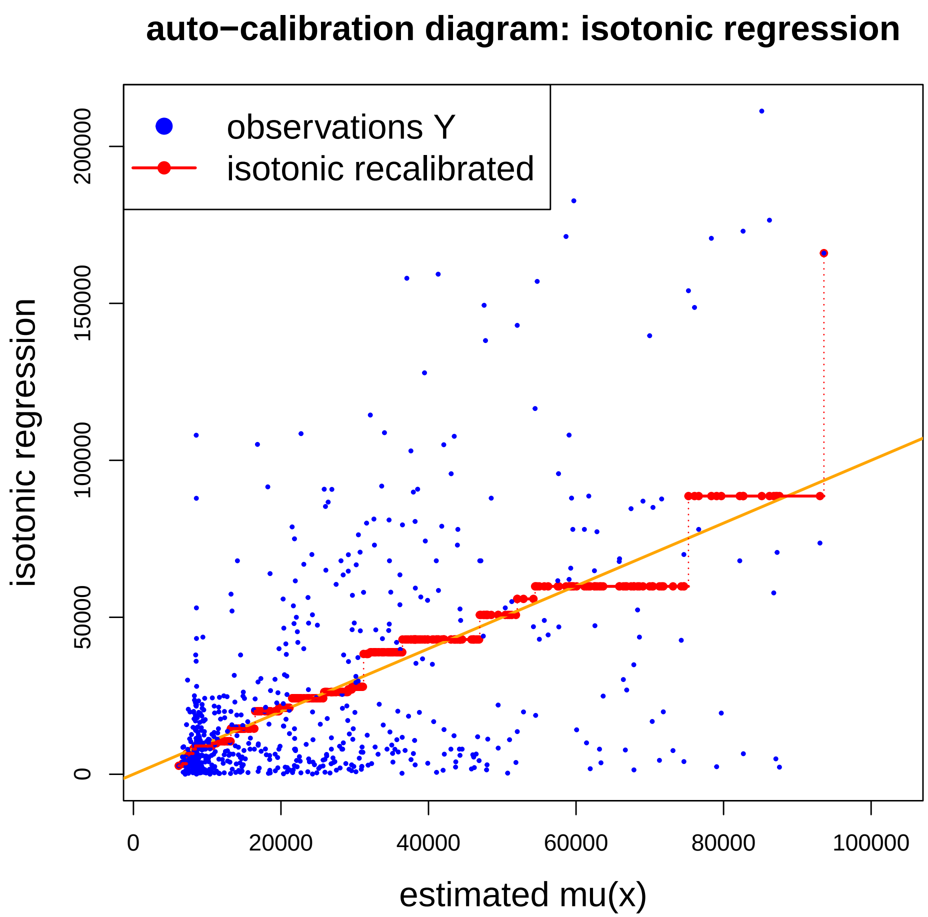
Isotonic regression solves the following optimization problem (for positive weights w_i)

$$\widehat{\boldsymbol{m}} = \underset{\boldsymbol{m}=(m_1,\ldots,m_n)^\top \in \mathbb{R}^n}{\arg\min} \sum_{i=1}^{n} w_i \left(Y_i - m_i\right)^2,$$

subject to $m_k \leq m_j \iff \mu(\boldsymbol{x}_k) \leq \mu(\boldsymbol{x}_j)$.

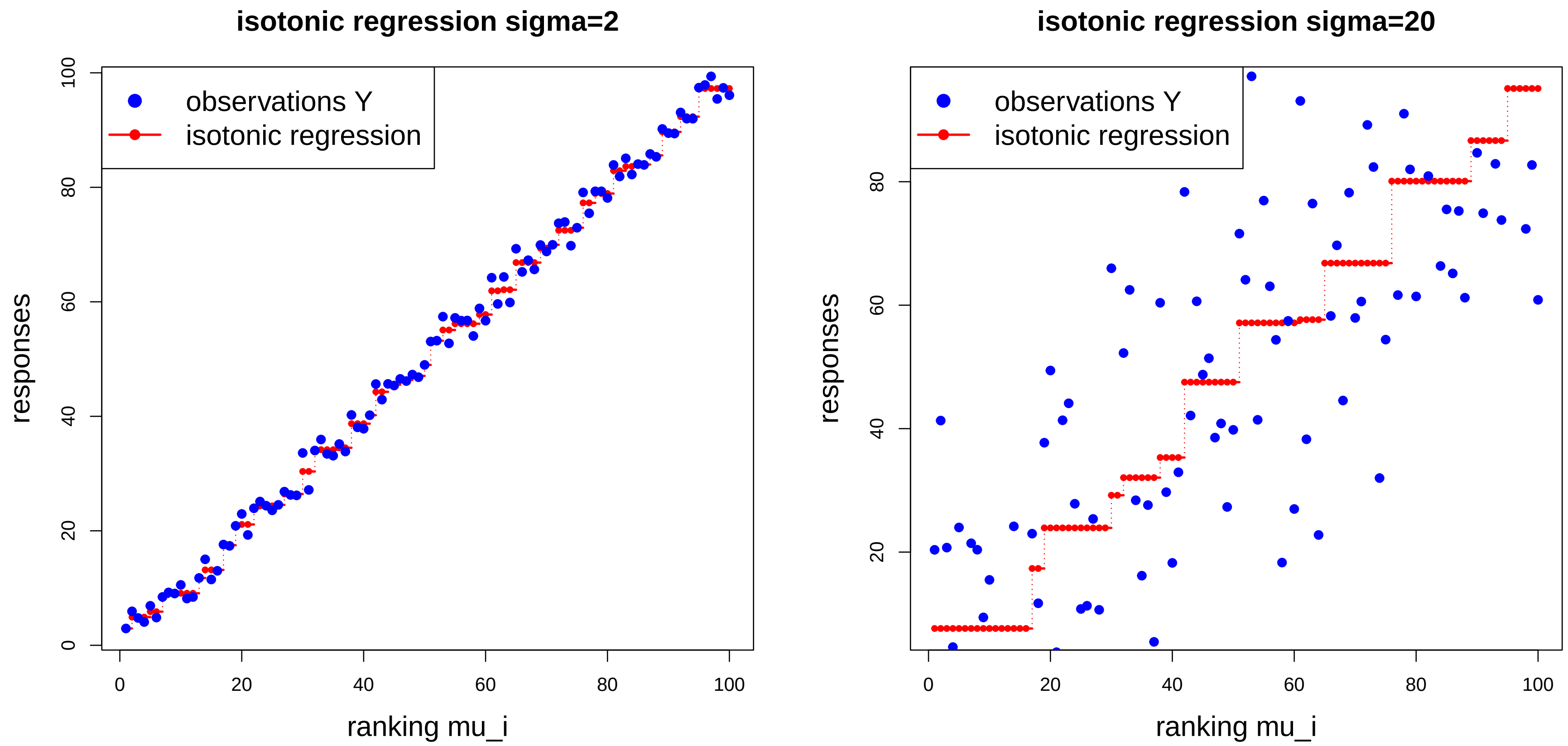
- Isotonic regression $(Y_i, \widehat{m}_i)_{i=1}^n = (Y_i, \widehat{m}(\boldsymbol{x}_i))_{i=1}^n$ is (empirically) auto-calibrated.
- Isotonic regression does not need any hyperparameter tuning.
- Isotonic regression is solved by the pool adjacent violators (PAV) algorithm; see Ayer et al. (1955).
- *Any* strictly consistent loss L for mean estimation gives the *same* solution $\widehat{\boldsymbol{m}}$.

Isotonic recalibration: step function



- Isotonic regression: natural (optimal) binning without hyperparameter tuning.
- May over-fit at both ends: merge smallest and largest bins.

Isotonic regression and signal-to-noise ratio



(left): high signal-to-noise ratio (right): low signal-to-noise ratio

Isotonic regression under low signal-to-noise ratio

- Insurance data typically has a low signal-to-noise ratio.

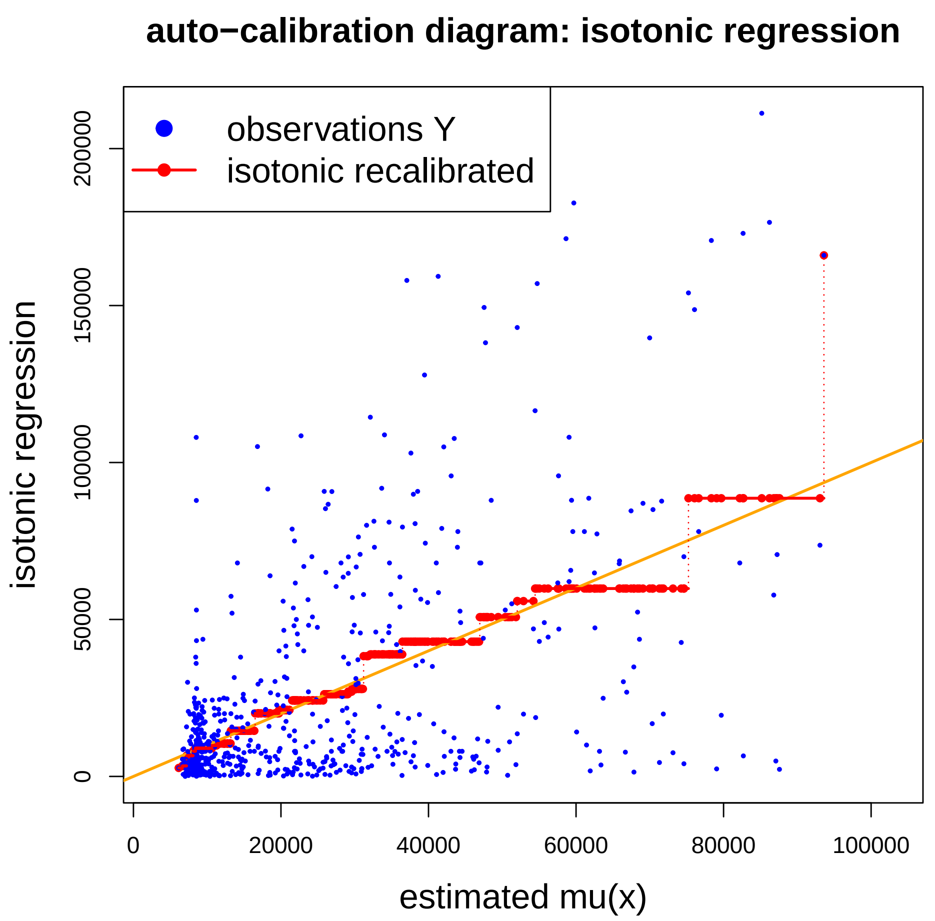
- **Theorem** (W.–Ziegel, 2023)**.** The expected number of steps in the isotonic (step) regression function is increasing in the signal-to-noise ratio.

- In low signal-to-noise ratio cases: isotonic regression leads to a low complexity partition of the covariate space

$$\mathcal{X}_k := \{\boldsymbol{x};\ \widehat{m}(\boldsymbol{x}) = m_k\},$$

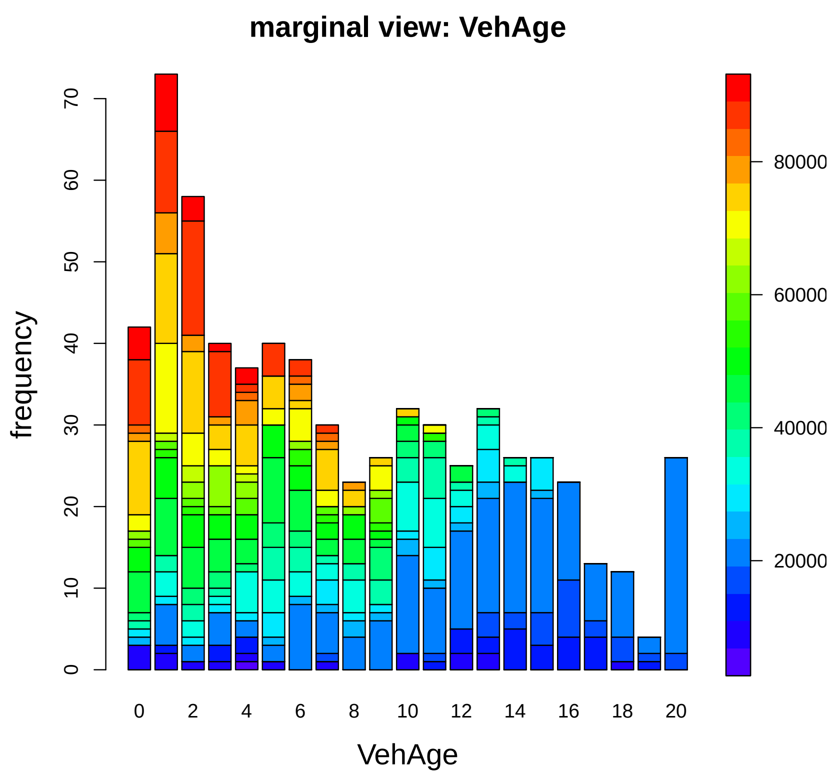
where m_k are the different values of the isotonic regression step function.

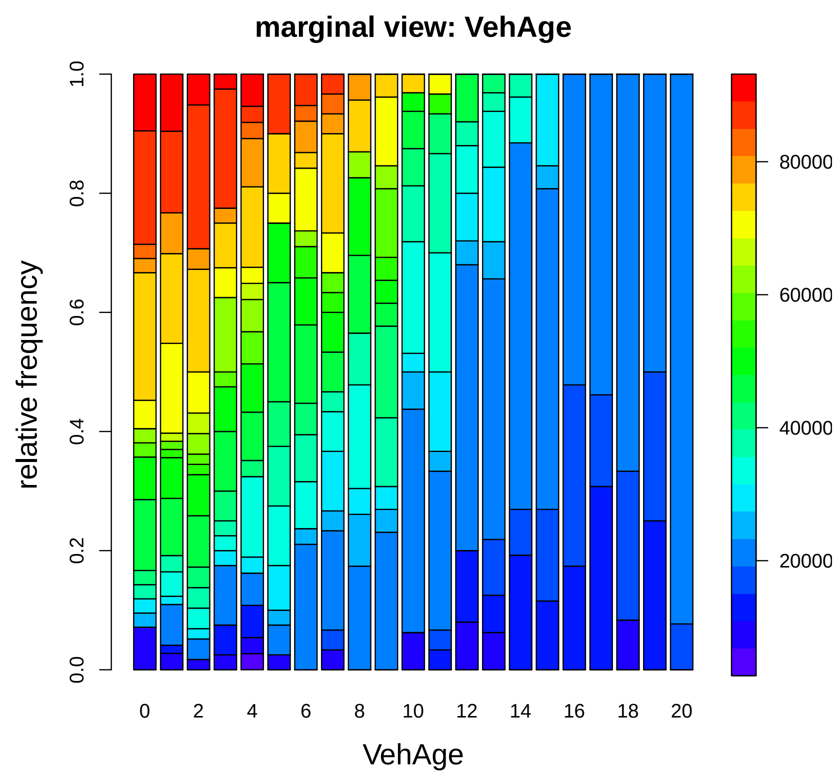
Isotonic recalibration: explainability



This neural network regression function results in 23 different values m_k of the isotonic regression function $\Rightarrow$ explainability through partition $(\mathcal{X}_k)_{k=1}^{23}$.

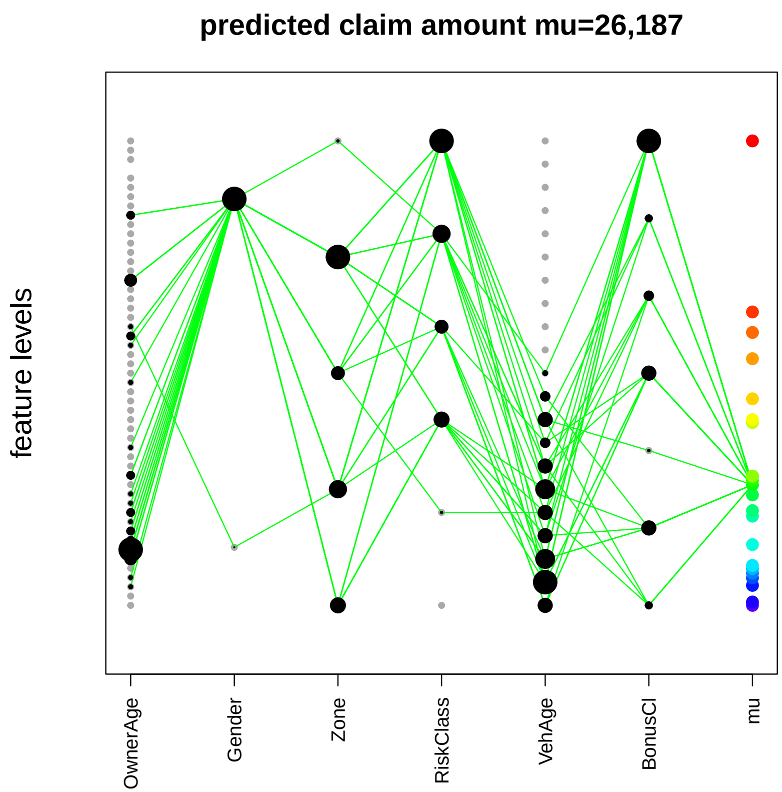
Attribution to individual covariate components

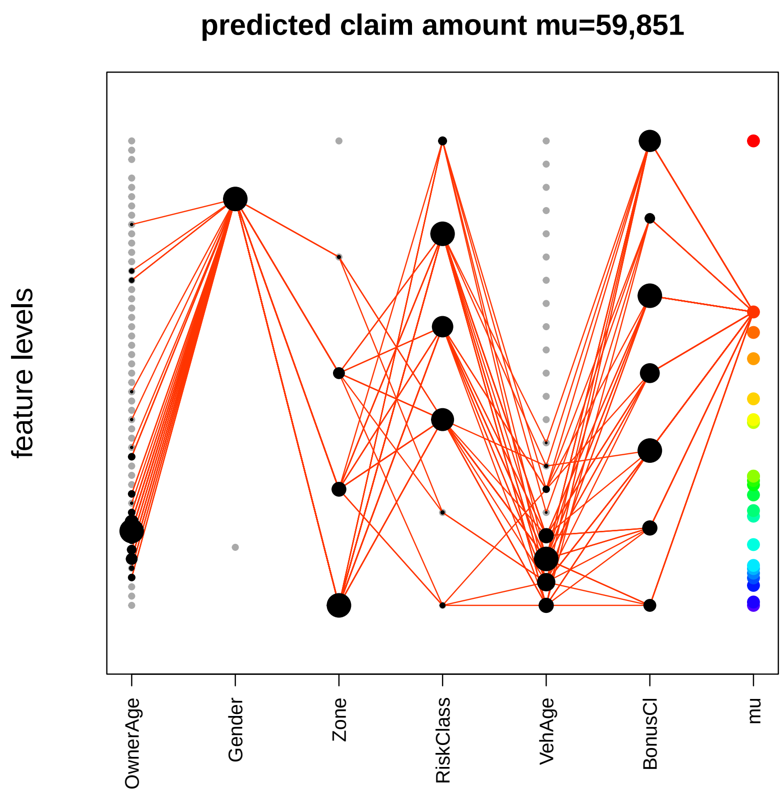




- Covariate space decomposition into 23 different values $(\mathcal{X}_k)_{k=1}^{23}$.
- No hyperparameters involved, only isotonic property is used.

Partition of feature space





- **Score decomposition**

More on auto-calibration

- The true regression function μ^* is auto-calibrated.

- There are (infinitely) many auto-calibrated regression functions:
  Choose a measurable function $\boldsymbol{x} \mapsto g(\boldsymbol{x})$, the following is auto-calibrated

$$\mu_g(\boldsymbol{x}) := \mathbb{E}\left[Y \middle| g(\boldsymbol{x})\right].$$

- In particular, the following is auto-calibrated

$$\mu_{\mathrm{rc}}(\boldsymbol{x}) := \mathbb{E}\left[Y \middle| \mu(\boldsymbol{x})\right]. \tag{1}$$

- The simplest auto-calibrated price is the homogeneous one $\mu_0 := \mathbb{E}[Y]$.

- For an increasing sequence of information sets $\mathcal{F}_t \subset \mathcal{F}_{t+1}$, we can generate an increasing set (in convex order) of auto-calibrated prices.
  This reflects a way of constructing a martingale.

Murphy's score decomposition

- Murphy (1973) proposed a score decomposition for analyzing (binary) meteorology forecasts, see also Semenovich–Dolman (2020). (Only) recently, this has been generalized to arbitrary responses by Pohle (2020) and Gneiting–Resin (2020).

- We have Murphy's score decomposition

$$\mathbb{E}\left[(Y-\mu(\boldsymbol{x}))^2\right] = \mathsf{UNC} - \mathsf{DSC} + \mathsf{MSC},$$

  with uncertainty, discrimination (resolution) and miscalibration, respectively,

$$\begin{aligned}
\mathsf{UNC} &= \mathbb{E}\left[(Y-\mathbb{E}[Y])^2\right] = \mathrm{Var}(Y) \geq 0,\\
\mathsf{DSC} &= \mathbb{E}\left[(Y-\mathbb{E}[Y])^2\right] - \mathbb{E}\left[(Y-\mu_{\mathrm{rc}}(\boldsymbol{x}))^2\right] \geq 0,\\
\mathsf{MSC} &= \mathbb{E}\left[(Y-\mu(\boldsymbol{x}))^2\right] - \mathbb{E}\left[(Y-\mu_{\mathrm{rc}}(\boldsymbol{x}))^2\right] \geq 0.
\end{aligned}$$

- This holds for any strictly consistent loss function L for mean estimation.

- **Conclusions**

Concluding remarks

- Estimation should be based on strictly consistent loss functions; Gneiting (2011).
- Bregman divergences are the only strictly consistent loss functions for mean estimation, Savage (1971) and Gneiting (2011).
- Any regression function should be auto-calibrated for insurance pricing.
- Isotonic recalibration restores the auto-calibration property (empirically).
- A low signal-to-noise ratio leads to a low complexity isotonic recalibrated function.
- Murphy's score decomposition gives a useful tool for model analysis.
- Model selection with Gini indices should only be done under auto-calibration.

References

[1] Ayer, M., Brunk, H.D., Ewing, G.M., Reid, W.T., Silverman, E. (1955). An empirical distribution function for sampling with incomplete information. *Annals of Mathematical Statistics* **26**, 641-647.

[2] Denuit, M., Charpentier, A., Trufin, J. (2021). Autocalibration and Tweedie-dominance for insurance pricing in machine learning. *Insurance: Mathematics & Economics* **101/B**, 485-497.

[3] Fissler, T., Lorentzen, C., Mayer, M. (2022). Model comparison and calibration assessment: user guide for consistent scoring functions in machine learning and actuarial practice. *arXiv*:2202.12780.

[4] Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association* **106/494**, 746-762.

[5] Gneiting, T., Ranjan, R. (2013). Combining predictive distributions. *Electronic Journal of Statistics* **7**, 1747-1782.

[6] Gneiting, T., Resin, J. (2022). Regression diagnostics meets forecst evaluation: conditional calibration, reliability diagrams, and coefficient of determination. *arXiv*:2108.032110v3.

[7] Hosmer, D.W., Lemeshow, S. (1980). Goodness of fit tests for the multiple logistic regression model. *Communications in Statistics - Theory and Methods* **9**, 1043-1069.

[8] Krüger, F., Ziegel, J.F. (2021). Generic conditions for forecast dominance. *Journal of Business & Economics Statistics* **39/4**, 972-983.

[9] Lindholm, M., Lindskog, F., Palmquist, J. (2023). Local bias adjustment, duration-weighted probabilities, and automatic construction of tariff cells. *Scandinavian Actuarial Journal*, to appear.

[10] Loader, C. (1999). *Local Regression and Likelihood*. Springer.

[11] Menon, A.K., Jiang, X., Vembu, S., Elkan, C., Ohno-Machado, L. (2012). Predicting accurate probabilities with ranking loss. *ICML'12: Proceedings of the 29th International Conference on Machine Learning*, 659-666.

[12] Murphy, A.H. (1973). A new vector partition of the probability score. *Journal of Applied Meteorology* **12/4**, 595-600.

[13] Pohle, M.-O. (2020). The Murphy decomposition and the calibration-resolution principle: A new per-spective on forecast evaluation. *arXiv*:2005.01835.

[14] Savage, L.J. (1971). Elicitable of personal probabilities and expectations. *Journal of the American Statistical Association* **66/336**, 783-810.

[15] Schervish, M.J. (1989). A general method of comparing probability assessors. *The Annals of Statistics* **17/4**, 1856-1879.

[16] Semenovich, D., Dolman, C. (2020). What makes a good forecast? Lessons from meteorology. *20/20 All-Actuaries Virtual Summit*, The Institute of Actuaries, Australia.

[17] Tasche, D. (2021). Calibrating sufficiently. *Statistics: A Journal of Theoretical and Applied Statistics* **55/6**, 1356-1386.

[18] Tsyplakov, A. (2013). Evaluation of probabilistic forecasts: proper scoring rules and moments. *SSRN Manuscript* ID 2236605.

[19] Wüthrich M.V. (2023). Model selection with Gini indices under auto-calibration. *European Actuarial Journal*, to appear.

[20] Wüthrich, M.V., Merz, M. (2023). *Statistical Foundations of Actuarial Learning and its Applications*. Springer Actuarial.

[21] Wüthrich, M.V., Ziegel, J. (2023). Isotonic recalibration under a low signal-to-noise ratio. *arXiv*:2301.0269.