Auto-Calibration and Isotonic Recalibration

Mario V. Wüthrich
RiskLab, ETH Zurich

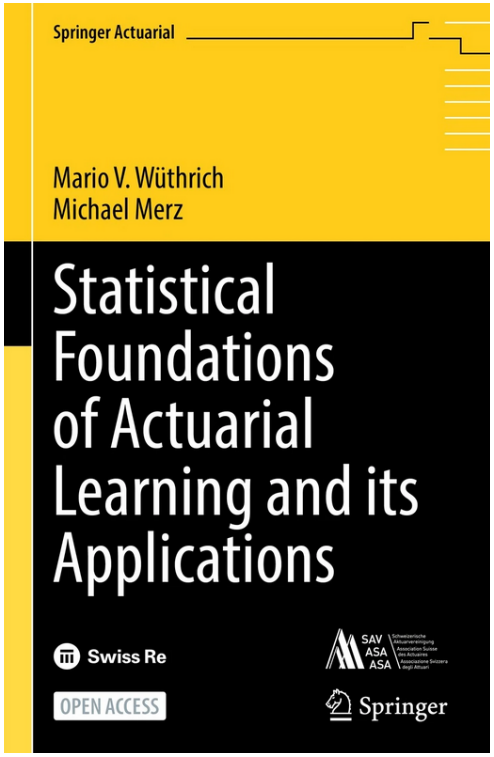
RiskLab Switzerland

Joint work with Johanna Ziegel

June 15, 2023
Insurance Data Science Conference
Bayes Business School, City, University of London

Overview

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



- **Introduction: regression models**

Best-estimate premium

- For known data generating model, compute true best-estimate of a policyholder with features $\boldsymbol{x}$ by

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

- For unknown data generating model, estimate μ^* from a sample $(Y_i, \boldsymbol{x}_i)_{i=1}^n$ that has been generated by this unknown model. Solve

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

  for a given model class $\mathcal{M}$, and for a strictly consistent loss function L for mean estimation; see Gneiting (2011).

- Popular model selection criteria:
  - ⋆ out-of-sample losses (under strictly consistent loss functions);
  - ⋆ auto-calibration and global balance property;
  - ⋆ conditional T-reliability diagrams and score decompositions.

- **Auto-calibration and isotonic recalibration**

Auto-calibration

- Literature on auto-calibration: Schervish (1989), Menon et al. (2012), Tsyplakov (2013), Gneiting–Ranjan (2013), Pohle (2020), Gneiting–Resin (2022); Tasche (2021), Krüger–Ziegel (2021); Denuit et al. (2021), Fissler et al. (2022), W. (2023), Lindholm et al. (2023), W.–Ziegel (2023), ...

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

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

▷ This means that every price cohort $\mu(\boldsymbol{x})$ is on 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

- **Binning** (Hosmer–Lemeshow (1980) χ^2-test, Lindholm et al. (2023)): Build disjoint price intervals (bins) $I_k = [a_k, a_{k+1})$ and consider the average claim in each bin

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

- **Local Regression** (Loader (1999), Denuit et al. (2021)): Binning is a discretized version of a local regression (or kernel smoother) that regresses the responses Y from the price cohorts $\mu(\boldsymbol{x})$

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

- Both methods are sensitive to hyperparameter selection.

Isotonic recalibration

- Isotonic regression is a non-parametric way to restore the auto-calibration property.

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

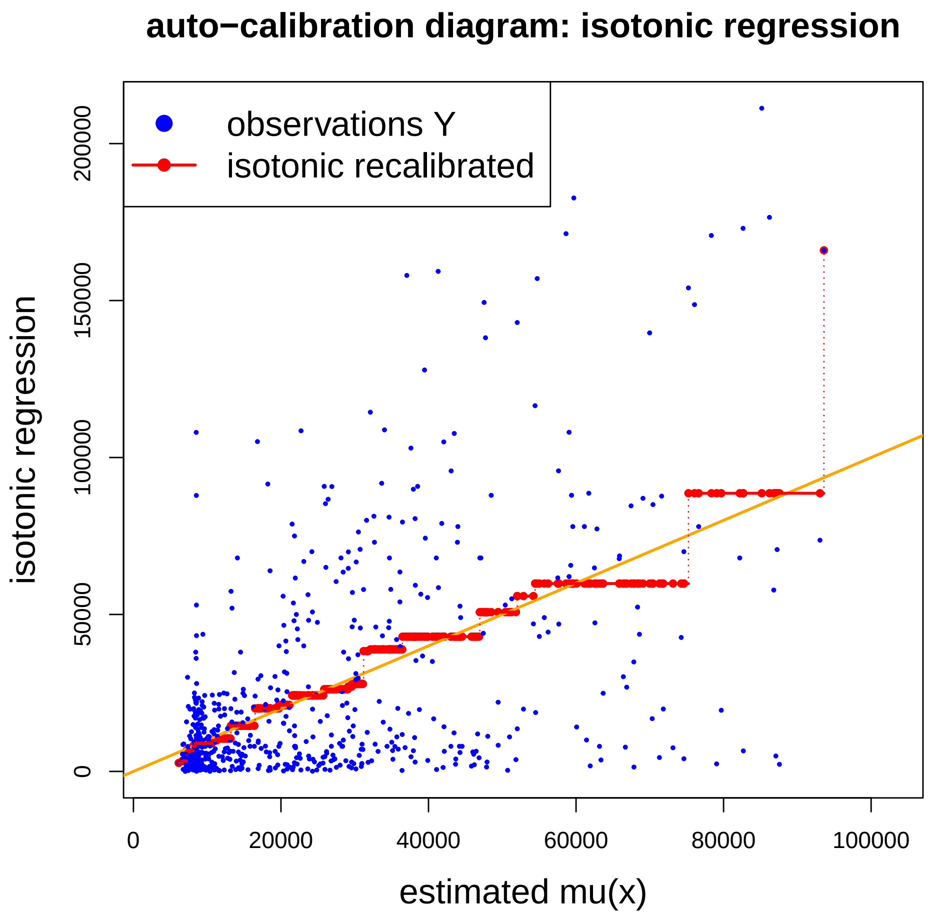
$$\widehat{\boldsymbol{m}} = \underset{\boldsymbol{m}=(m_1,\ldots,m_n)^\top \in \mathbb{R}^n}{\arg\min} \sum_{i=1}^{n} w_i \left(Y_i - m_i\right)^2,$$

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.

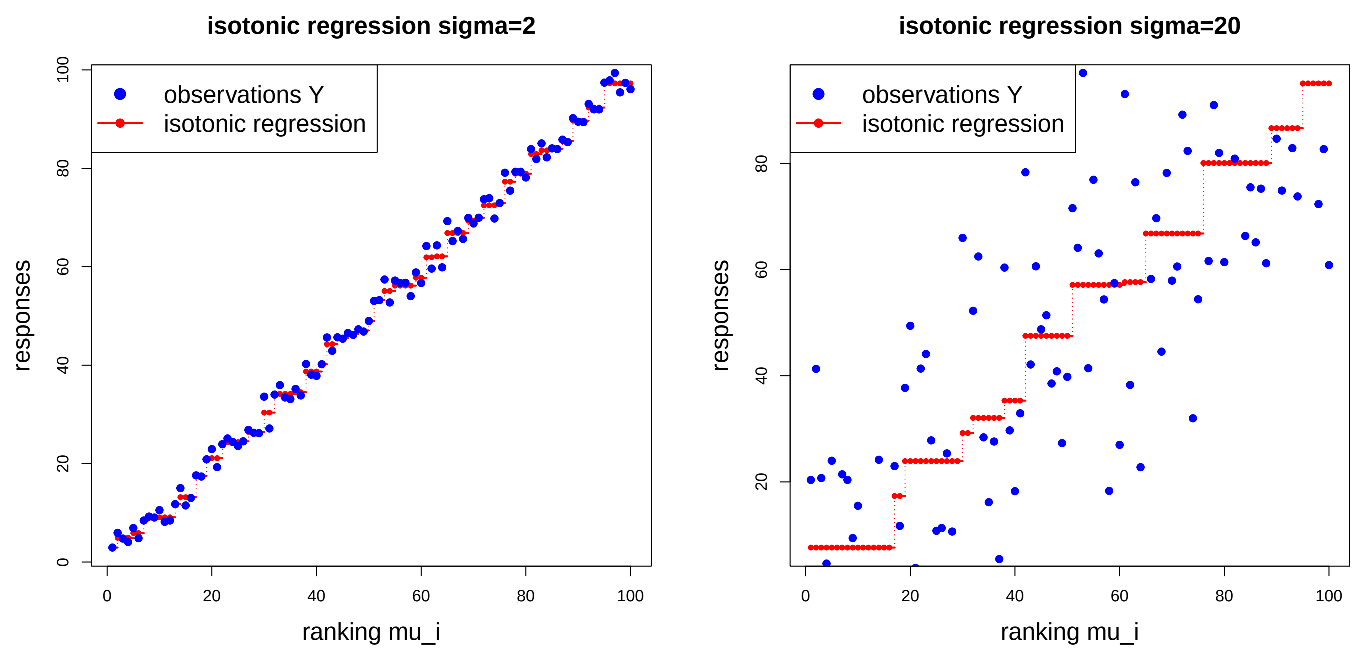
- Isotonic regression $(Y_i, \widehat{m}_i)_{i=1}^n = (Y_i, \widehat{m}(\boldsymbol{x}_i))_{i=1}^n$ is (empirically) auto-calibrated.

Isotonic recalibration: step function



- Isotonic regression: natural (optimal) binning without hyperparameter tuning.

Isotonic regression and signal-to-noise ratio



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

- Low signal-to-noise ratio: isotonic regression leads to low complexity partition of the feature space $\Rightarrow$ explainability.

- **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 functions.
- A low complexity partition of the feature space gives explainability.

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 perspective 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* **13/1**, 469-477.

[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.