Local bias adjustment, duration-weighted probabilities, and automatic construction of tariff cells

Mathias Lindholm
Stockholm University

IDS, June, 2023

The presentation is based on the paper

"*Local bias adjustment, duration-weighted probabilities, and automatic construction of tariff cells*"

in Scandinavian Actuarial Journal (available online), written together with Filip Lindskog and Johan Palmquist

I will also refer to results from a recent SSRN pre-print on effects of violating GLM assumptions, written together with Taariq Nazar

Outline of the presentation

- Background to predictive non-life insurance pricing
- Bias, auto-calibration, and auto-tariffication
- Numerical examples

Remarks.

- Details about dispersion modelling have been omitted, although included in the numerical examples (see appendix)
- Notation will at times be informal

Non-life insurance pricing

- What is it that we observe?
- Data (Z, X, W), where
  - $Z \in \mathbb{R}_+$ is the response, e.g. no. of claims or claim cost
  - $X \in \mathbb{X}$ is a covariate vector
  - $W \in \mathbb{R}_+$ is a weight, e.g. duration

For the remainder of the presentation we will refer to W as duration

Non-life insurance pricing

- In practice we are often interested in e.g.
  - claim *frequency*
  - *average* claim cost
  - the pure premium = claim cost / duration

  and we model variables of the type $Y = Z/W$
- In retrospect W is known, but when a premium is paid W is random
- Costs and premium calculations?

Non-life insurance pricing

- Let Z denote the claim cost
- The *actuarially fair premium*, $\pi(X)$, is defined by

$$\mathbb{E}[W\pi(X) \mid X] = \mathbb{E}[Z \mid X] \quad (= \mathbb{E}[WY \mid X])$$

$\Leftrightarrow$ Expected earned premium $=$ expected claim cost

- That is

$$\pi(X) := \frac{\mathbb{E}[Z \mid X]}{\mathbb{E}[W \mid X]} = \mathbb{E}\left[\underbrace{\frac{W}{\mathbb{E}[W \mid X]}}_{=\text{duration weights}=:\mathbb{P}_W} Y \mid X\right] =: \mathbb{E}_{\mathbb{P}_W}[Y \mid X],$$

and $\pi(X)$ is the *expected duration adjusted claim cost*

Non-life insurance pricing

- Next, note that if we define

$$\mu_W(X) := \mathbb{E}_{\mathbb{P}_W}[Y \mid X],$$

it holds that

$$\mu_W(X) \in \arg\min_g \mathbb{E}_{\mathbb{P}_W}[(Y - g(X))^2],$$

for all g such that $\mathbb{E}[g(X)^2] < \infty$

- Thus, by using the rescaling $\mathbb{P}_W$ we can optimise "as usual"

Non-life insurance pricing

Remarks.

- If we make the additional "GLM" *assumption*

$$\mathbb{E}[Z \mid X, W] = W\lambda(X),$$

  it follows that

$$\mathbb{E}_{\mathbb{P}_W}[Y \mid X] = \mathbb{E}[Y \mid X],$$

  see (Lindholm et al. 2023, Prop. 2.1)
- GLMs, or EDF models, can be treated analogously as above by replacing the L^2 unit deviance, see Lindholm et al. (2023)

Bias and auto-calibration

- Assume that we have a predictor $\widehat{\mu}(X)$
- We will stay in L^2 and we say that this predictor is "good" if $\mathbb{E}_{\mathbb{P}_W}[(Y - \widehat{\mu}(X))^2]$ is small

What if we want to improve on $\widehat{\mu}$?

Bias and auto-calibration

Proposition 1

The following inequalities hold:

$$\mathbb{E}_{\mathbb{P}_W}[(Y - \widehat{\mu}(X))^2] \geq \mathbb{E}_{\mathbb{P}_W}[(Y - \mathbb{E}_{\mathbb{P}_W}[Y \mid \widehat{\mu}(X)])^2] \geq \mathbb{E}_{\mathbb{P}_W}[(Y - \mu_W(X))^2]$$

Note that

- the first inequality in the proposition is an equality iff

$$\widehat{\mu}(X) = \mathbb{E}_{\mathbb{P}_W}[Y \mid \widehat{\mu}(X)]$$

  which corresponds to that $\widehat{\mu}(X)$ is *auto-calibrated*, see (Krüger & Ziegel 2021, Def. 3.1) and (Denuit et al. 2021, Sec. 5.1)
- if we believe $\widehat{\mu}(X)$ is close to $\mu_W(X)$ we do not expect much improvement (we'll come back to this in the examples)

Bias and auto-calibration

- Further, Prop. 1 tells us that

$$\widehat{\mu}^*_W(X) := \mathbb{E}_{\mathbb{P}_W}[Y \mid \widehat{\mu}(X)]$$

is the natural candidate for an improved predictor
- This predictor is not attainable based on a finite sample
- Suggestion: **Partition based on $\widehat{\mu}$!** That is,
  - Since $\widehat{\mu} \in \mathbb{R}_+$, split $\mathbb{R}_+$ into κ bins (somehow) according to $b_0 = 0 < b_1 < \ldots < b_{\kappa-1} < b_\kappa = +\infty$
  - Using $\widehat{\mu}$ create a partition of the covariate space according to

$$B_k := \{x \in \mathbb{X} : \widehat{\mu}(x) \in [b_{k-1}, b_k)\}, \quad \mathbb{X} = \cup_{k=1}^{\kappa} B_k$$

  - Introduce the piecewise constant bias adjusted predictor

$$\widehat{\mu}^{\text{ba}}_W(X) := \sum_{k=1}^{\kappa} \mathbb{E}_{\mathbb{P}_W}[Y \mid X \in B_k] 1_{\{X \in B_k\}}$$

which is auto-calibrated, see Lindholm et al. (2023)

Bias and auto-calibration

The bias adjusted predictor to be used in practice is the following plug-in predictor of the L^2 minimiser

$$\widehat{\widehat{\mu}^{\text{ba}}_{W}}(x) := \sum_{k=1}^{\kappa} \widehat{\mathbb{E}}_{\mathbb{P}_W}[Y \mid X \in B_k] \mathbb{1}_{\{X \in B_k\}}$$

$$= \sum_{k=1}^{\kappa} \frac{\sum_{i=1}^{m} w_i y_i \mathbb{1}_{B_k}(x_i)}{\sum_{i=1}^{m} w_i \mathbb{1}_{\{X \in B_k\}}} \mathbb{1}_{\{x \in B_k\}}$$

Bias and auto-calibration

- How to choose the B_ks (or rather the b_ks)?
- Suggestion: Minimise MSEP based on
  - equal duration binning, i.e. sort $(y_i, \widehat{\mu}(x_i), w_i)_i$ based on $\widehat{\mu}$ and split into equal duration bins
  - a duration weighted L^2-regression tree
  - ...

Note:

- We let data decide on the effective number of bins (or tariff cells) using MSEP
- This gives us a *data driven automatic construction of tariffs*

Numerical examples

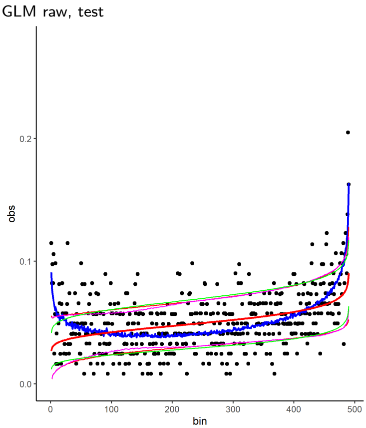
We consider two sets of data:

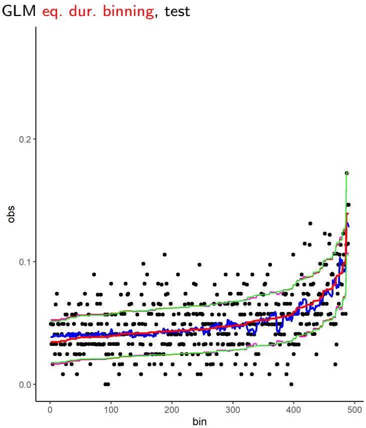
- Simulated Poisson data
- CASdatasets: `freMTPL`

The simulated data set is made to resemble the `freMTPL` data, in short:

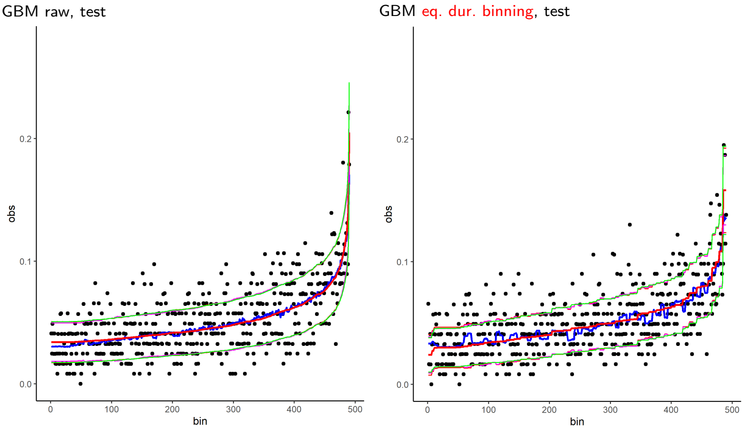
- $n = 300\,000$ policies, 80% used for training 20% used for out-of-sample test
- $\mathbb{E}[Z] = 0.05 =$ expected no. of claims for a single contract
- $\text{Var}(\mu(X)) \approx 0.02$

Simulated data





The bins on the x-axis correspond to the risk ordering based on the initial predictor $\widehat{\mu}$; model prediction (**thick red**); initial std. dev of Z (thin red); adjusted (magenta); Pois-Z ref. (green); truth (blue)



The bins on the x-axis correspond to the risk ordering based on the initial predictor $\widehat{\mu}$; model prediction (**thick red**); initial std. dev of Z (thin red); adjusted (magenta); Pois-Z ref. (green); truth (blue)

Simulated data

Conclusions

- A misspecified model can be corrected satisfactory
- A reasonably well specified model is not damaged
- Equal duration binning only use $\sim$ 100 bins

CASdatasets: freMTPL

Simulated data: No duration effects

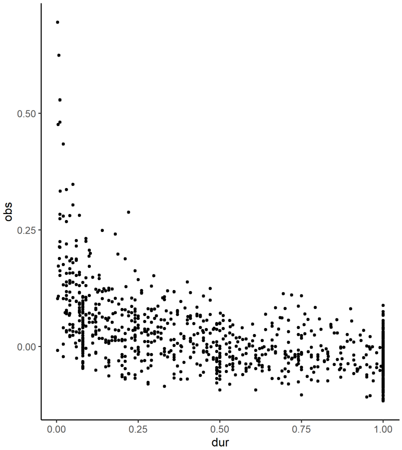
Real data: Potential duration effects

Modelling assumption, initial predictor:

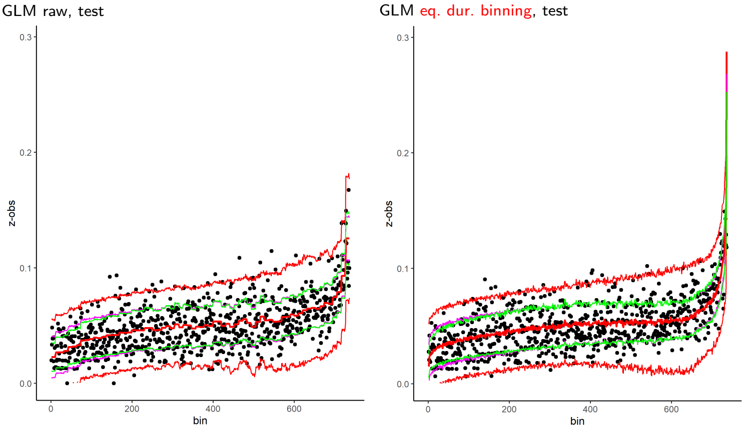
$\mathbb{E}[Z \mid X, W]$ and $\mathrm{Var}(Z \mid X, W)$ linear in W — "GLM"

Note: if this assumption is wrong, the plug-in estimator of $\mathrm{Var}(Z \mid X)$ will *be systematically over-estimated!*

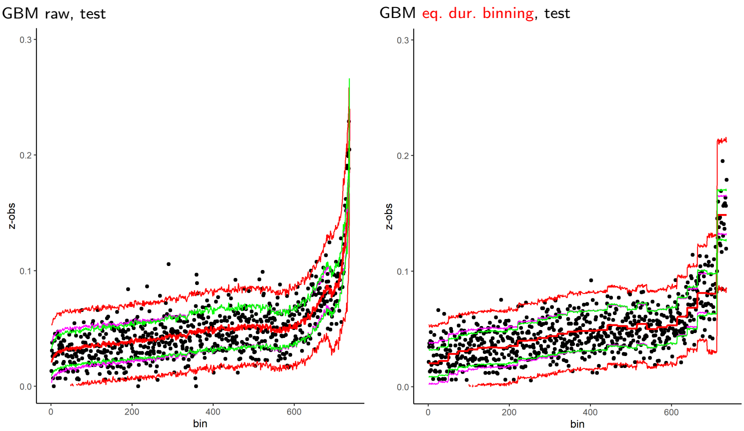
See, Lemma 3.3 in Lindholm & Nazar (2023)



- If the GLM moment assumption is satisfied, the points should be evenly scattered around 0
- Here, low duration seems to imply higher risk



The bins on the x-axis correspond to the risk ordering based on the initial predictor $\widehat{\mu}$; model prediction (**thick red**); initial std. dev of Z (thin red); adjusted (magenta); Pois-Z ref. (green)



The bins on the x-axis correspond to the risk ordering based on the initial predictor $\widehat{\mu}$; model prediction (**thick red**); initial std. dev of Z (thin red); adjusted (magenta); Pois-Z ref. (green)

CASdatasets: freMTPL

Conclusions

- A misspecified model can be corrected satisfactory
- Duration weights matter – GLM weights not entirely correct
- Bias regularising the variance makes the model very close to Poisson
- Equal duration binning only use $\sim$ 100 bins

Summary and conclusions

- Bias adjust!
- The piecewise constant technique is simple!
- Data decides the size of the tariff!
- Duration weights matter!

For more on effects of violating GLM assumptions, see Lindholm & Nazar (2023)

Thank you for your attention!

References I

Denuit, M., Charpentier, A. & Trufin, J. (2021), 'Autocalibration and Tweedie-dominance for insurance pricing with machine learning', Insurance: Mathematics and Economics **101**, 485–497.

Krüger, F. & Ziegel, J. F. (2021), 'Generic conditions for forecast dominance', Journal of Business & Economic Statistics **39**(4), 972–983.

Lindholm, M., Lindskog, F. & Palmquist, J. (2023), 'Local bias adjustment, duration-weighted probabilities, and automatic construction of tariff cells', Scandinavian Actuarial Journal (available online) .

Lindholm, M. & Nazar, T. (2023), 'On duration effects in non-life insurance pricing', Available at SSRN 4474908 .

Wüthrich, M. V. & Ziegel, J. (2023), 'Isotonic recalibration under a low signal-to-noise ratio', arXiv preprint arXiv:2301.02692 .

Appendix

A bit more details and related methods

- Dispersion modelling = *variance* modelling

  How to do this?

  - "GLM" assumption – the variance of $Z \mid X, W$ is linear in W (as in the numerical illustrations)

    $\Rightarrow$ variance decomposition!
  - estimate the functional form of $\mathbb{E}[W \mid X]$, i.e.

$$\widehat{\nu} \in \arg\min_{g} \mathbb{E}_{\mathbb{P}_W}[((Z - \widehat{\mathbb{E}}[W \mid X]\widehat{\mu}_W(X))^2 - g(X))^2]$$

A bit more details and related methods

Size of the tariff!?

- The performance of the original predictor matters! If you use an intercept only model, you cannot learn anything new

  $\Rightarrow$ risk-ordering of the original predictor!

  $\Rightarrow$ *isotonic regression*, see Wüthrich & Ziegel (2023), assumes a correct risk-ordering of the original predictor
- Effects of variation in the data generating process!?

  Result for isotonic regression (Wüthrich & Ziegel (2023)): *lowering the signal to noise ratio in the data generating process will reduce the size of the resulting tariff*

  (This is believed to hold for the current method as well)