On Duration Effects in Non-Life Insurance Pricing

Mathias Lindholm
Stockholm University

IDSC 18 June 2024

The presentation is based on the paper

On Duration Effects in Non-Life Insurance Pricing

written togheter with Taariq Nazar (Stockholm University), to appear in the European Actuarial Journal

Background

- Data consists of triplets (Z, X, W), where
  - Z corresponds to the response e.g. claim amount,
  - X is a vector of covariates,
  - W is an exposure weight e.g. policy duration
- Standard GLM assumptions

$$\mathbb{E}[Z \mid X, W] = W\mu(X), \text{ and } \mathrm{Var}[Z \mid X, W] = W\sigma^2(X) \tag{A1}$$

- ...more specificly: Tweedie

$$\mathrm{Var}[Z \mid X, W] = W\varphi\mu(X)^{\xi} \tag{A2}$$

  where "φ" is the dispersion parameter
- Note that both (A1) and (A2) are linear in W

Background

Remarks

- Parameter estimation: the influence of W under (A1) is not obvious
- Under (A2): hard to estimate φ if $\mu(X)$ is badly specified (e.g. too rigid)

Background estimation

- Estimation of μ typically based on deviance loss functions, where the minimum is attained by the maximum likelihood estimator (MLE)
- We consider **Bregman deviance losses**
(e.g. Poisson, and Gamma with fixed dispersion, under (A1))
- The deviance function for $Y = Z/W$ can be written as

$$D_{\text{Breg}}(Y, \mu) \propto W d(Y, \mu),$$

where $d(Y, \mu)$ is the so-called unit deviance function, see e.g. Ohlsson & Johansson (2010), Wüthrich & Merz (2023)

Background estimation

- As discussed in Lindholm et al. (2023), the MLE for μ corresponds to the **empirical version** of the minimiser

$$\pi(X) \in \operatorname{argmin}_f \mathbb{E}[Wd(Y, f(X))], \tag{1}$$

  where the minimisation is over suitable X-measurable functions f (see ref. for details)
- As shown in Lindholm et al. (2023), the population minimiser $\pi(X)$ is given by

$$\pi(X) = \frac{\mathbb{E}[Z \mid X]}{\mathbb{E}[W \mid X]}, \tag{2}$$

  which does not assume (A1) and does not rely on any specific assumptions regarding the dependence between, Z, X, and W

Background estimation

Remarks

- From (2) it is clear that $\pi(X)$ will differ from $\mu(X)$ unless assumption (A1) is satisfied
- Given a reasonably well specified model, the above suggests that the MLE will be a consistent estimator of $\pi(X)$, which may, or may not, coincide with $\mu(X)$
- Note that $\pi(X)$ corresponds to the duration adjusted actuarially fair premium, since $\pi(X)$ satisfies the relation

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

Background estimation

Note that

- Consistency relies on that we know that the functional form of the true model is a GLM (or some other model class)
- In practice the true model is unknown, and a misspecified model for $\mathbb{E}[Z \mid X]$ will lead to local bias, see e.g. Lindholm et al. (2023) and Wüthrich & Ziegel (2023)
- Local bias will contaminate estimation of the dispersion parameter φ

Estimators and asymptotics

Estimators and asymptotics

- Focus will be on estimating the mean function for a specific covariate vector $X = x$, without assuming any specific functional form of $\mu(X)$
- Given sufficiently many observations of $X = x$, we may estimate $\mu(x)$ as a parameter
- This is a reasonable assumption when we consider the situation of letting the sample size tend to infinity

Estimators and asymptotics

Proposition 1

Consider an i.i.d. sample $(Z_i, X_i, W_i)_{i=1}^m = (Z_i, x, W_i)_{i=1}^m$ *and define* $Y_i := Z_i / W_i$. *The estimator* $\widehat{\mu}_m(x)$ *that minimises the duration weighted Bregman deviance is given by*

$$\widehat{\mu}_m(x) = \frac{\widehat{\mathbb{E}}_m[Z \mid X = x]}{\widehat{\mathbb{E}}_m[W \mid X = x]},$$

where

$$\widehat{\mathbb{E}}_m[Z \mid X = x] := \frac{1}{m} \sum_{i=1}^m Z_i, \quad \text{and} \quad \widehat{\mathbb{E}}_m[W \mid X = x] := \frac{1}{m} \sum_{i=1}^m W_i,$$

for which it holds that

$$\widehat{\mu}_m(x) \xrightarrow{p} \frac{\mathbb{E}[Z \mid X = x]}{\mathbb{E}[W \mid X = x]}, \quad \text{as } m \to \infty.$$

Estimators and asymptotics

Remarks (more in the paper)

- Proposition 1 does not assume
  - independence between Z and W
  - that the true data belongs to an EDF with expectation and variance being linear in W
- The predictor $\widehat{\mu}_m$ in Proposition 1 is always asymptotically actuarially fair in the sense of π from (2)

  ...but π is not guaranteed to equal μ unless (A1) hold!

Estimators and asymptotics

Dispersion modelling

- Above we have seen that $\mathbb{E}[Z \mid X]$ and $\mathbb{E}[W \mid X]$ appeared as limiting objects
- When discussing dispersion and variation we will encounter

$$\mathrm{Var}[Z \mid X] = \mathbb{E}[\mathrm{Var}[Z \mid X, W]] + \mathrm{Var}[\mathbb{E}[Z \mid X, W]],$$

  or the corresponding expressions under Tweedie assumptions
- We will focus on **Pearson estimators** of φ:

$$\widehat{\varphi}_m^{\mathrm{P}}(x) := \frac{1}{m-1} \sum_{i=1}^{m} \frac{W_i (Y_i - \widehat{\mu}_m(x))^2}{\widehat{\mu}_m^{\xi}(x)} \qquad (3)$$

  ...since there is trouble with consistency of deviance based estimators, see Lindholm & Nazar (2024)

Estimators and asymptotics

Proposition 2

Given an i.i.d. sample $(Z_i, X_i, W_i)_{i=1}^m = (Z_i, x, W_i)_{i=1}^m$ *it holds that*

$$\widehat{\varphi}_m^{\mathrm{P}}(x) \xrightarrow{p} \varphi^{*,\mathrm{P}}(x) = \overline{\varphi}(x) - \frac{\mathbb{E}[W \mid X = x]^{\xi-1}}{\mathbb{E}[Z \mid X = x]^{\xi}} \operatorname{Cov}\left[\frac{Z^2}{W}, W \;\middle|\; X = x\right],$$

as $m \to \infty$, *where*

$$\overline{\varphi}(x) := \frac{\mathbb{E}[W \mid X = x]^{\xi-1} \operatorname{Var}[Z \mid X = x]}{\mathbb{E}[Z \mid X = x]^{\xi}}. \tag{4}$$

If the underlying data generating process agrees with moment assumptions (A1) *and* (A2) *then* $\varphi^{*,\mathrm{P}}(x) = \varphi(x)$ *and* $\overline{\varphi}(x) > \varphi(x)$.

Estimators and asymptotics

Remarks

- Observing $\widehat{\varphi}^{*,\mathrm{P}}(x) \geq \widehat{\overline{\varphi}}(x)$ indicates violation of (A1) and (A2), see Lindholm & Nazar (2024)

$\rightarrow$ The same conclusions hold true if we observe $\widehat{\varphi}^{*,\mathrm{P}} \geq \widehat{\overline{\varphi}}(x)$

- The plug-in variance of Z based on $\widehat{\varphi}^{*,\mathrm{P}}(x)$ is only guaranteed to be consistent under (A1) and (A2), see Lindholm & Nazar (2024)

$\rightarrow$ An alternative to a plug-in variance estimator:

$$\widehat{\mathrm{Var}}_m[Z \mid X = x] := \frac{1}{m-1}\sum_{i=1}^{m}(Z_i - \underbrace{\widehat{\mathbb{E}}_m[W \mid X = x]\widehat{\mu}_m(x)}_{\widehat{\mathbb{E}}_m[Z|X=x]})^2 \tag{5}$$

which is consistent without assuming (A1) **and** (A2)**!**

Numerical illustrations

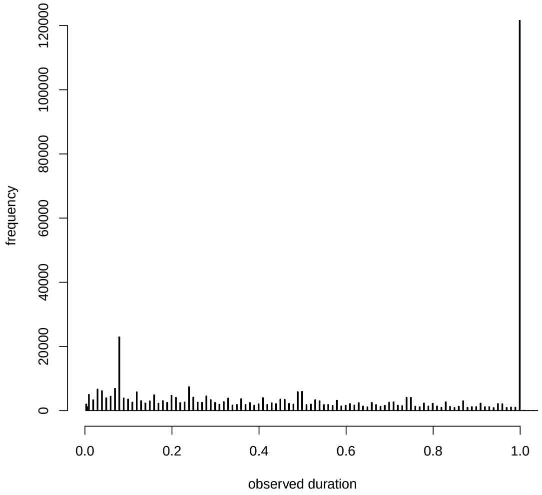
Setup

- Real insurance data: `freMTPLfreq`, see `CASdatasets`, and see Lindholm et al. (2023) for other examples
- Will use a Poisson GBM-model with linear weights in W, all standard parameters except tree depth, which is set to 2
- Optimal number of trees: 192

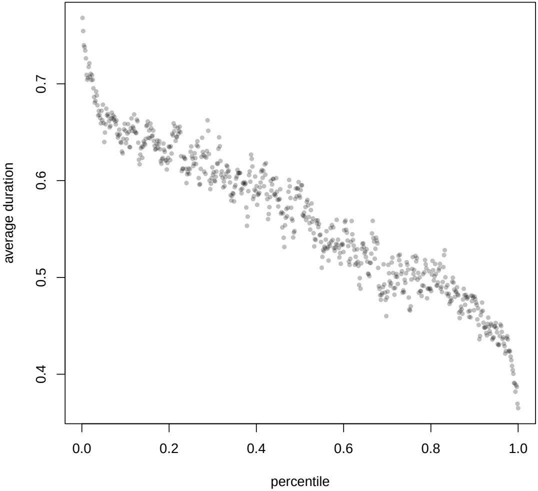
Model evaluation

- Use the $(\widehat{\mu}(x_i))_{i=1}^n$-predictions from the GBM and order these
- This gives us ordered $x_{(i)}$:s such that $\widehat{\mu}(x_{(i)})$ corresponds to the ith largest prediction
- The ordered data set is the split into $k = 200$ equally sized bins used to evaluate local performance

Numerical illustrations



Numerical illustrations



Numerical illustrations

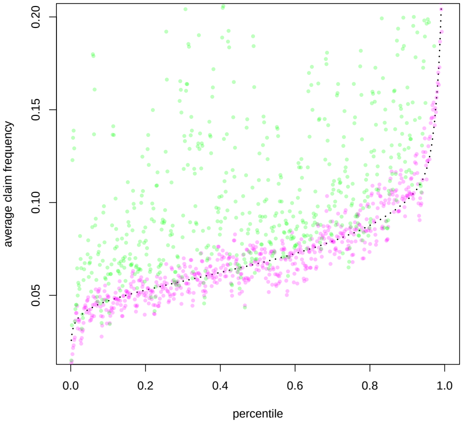


Figure: Dotted line: $\widehat{\mu}(x)$; Purple dots: $\widehat{\pi}(x) := \widehat{\mathbb{E}}[Z \mid X = x]/\widehat{\mathbb{E}}[W \mid X = x]$; Green dots: $\widehat{\mathbb{E}}[Y \mid X = x]$

Numerical illustrations

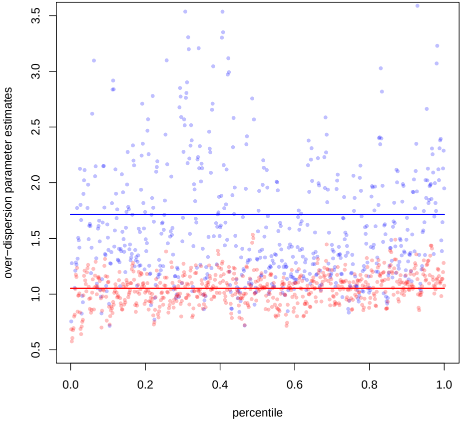


Figure: Blue dots: $\widehat{\varphi}^{\mathrm{P}}(x)$; Blue line: $\widehat{\varphi}^{\mathrm{P}}$; Red dots: $\widehat{\widehat{\varphi}}(x)$; Red line: average of the $\widehat{\widehat{\varphi}}(x)$s

Numerical illustrations

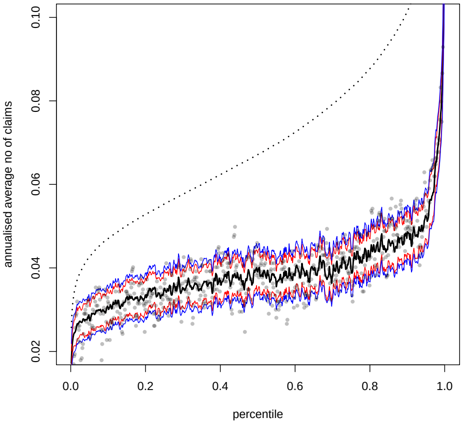


Figure: Black solid line: duration adjusted plug-in est. of $\mathbb{E}[Z \mid X = x]$; Grey dots: $\widehat{\widehat{\mathbb{E}}}[Z \mid X = x]$; Blue lines: plug-in est. of $\sqrt{\mathrm{Var}[Z \mid X = x]}$; Red lines: $\sqrt{\widehat{\widehat{\mathrm{Var}}}[Z \mid X = x]}$; Dashed line: $\widehat{\mu}(x)$

Numerical illustrations

Conclusions

- The dependence between Z and W matters in the bias calculations – supported by real data
- In the real data example the plug-in standard deviation variance **is on average 30% larger** than the corresponding local sample standard deviation using (5)
- This can be compared with that the using $\sqrt{\widehat{\varphi}^{\mathrm{P}}} \approx \sqrt{1.70} \approx 1.30$ instead of $\widehat{\widehat{\varphi}} = \widehat{\mathrm{Var}}[Z \mid X = x]/\widehat{\mathbb{E}}[Z \mid X = x] \approx 1.05$
- **Analysis "Trick"**: use the original predictor for risk ordering, and use simple sample variance estimators locally

References I

Lindholm, M., Lindskog, F. & Palmquist, J. (2023), 'Local bias adjustment, duration-weighted probabilities, and automatic construction of tariff cells', *Scandinavian Actuarial Journal* **2023**(10), 946–973.

Lindholm, M. & Nazar, T. (2024), 'On duration effects in non-life insurance pricing', *European Actuarial Journal*, to appear .

Ohlsson, E. & Johansson, B. (2010), *Non-life insurance pricing with generalized linear models*, Vol. 174, Springer.

Wüthrich, M. V. & Merz, M. (2023), *Statistical foundations of actuarial learning and its applications*, Springer Nature.

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