A machine learning approach based on survival analysis for IBNR frequencies in non-life reserving

Emil Hofman
emho@abgroup.com
Research Assistant / Actuary
University of Copenhagen / Alm Brand Group

Gabriele Pittarello
gabriele.pittarello@uniroma1.it
Postdoctoral Fellow
University of Torino

Munir Hiabu
mh@math.ku.dk
Assistant Professor
University of Copenhagen

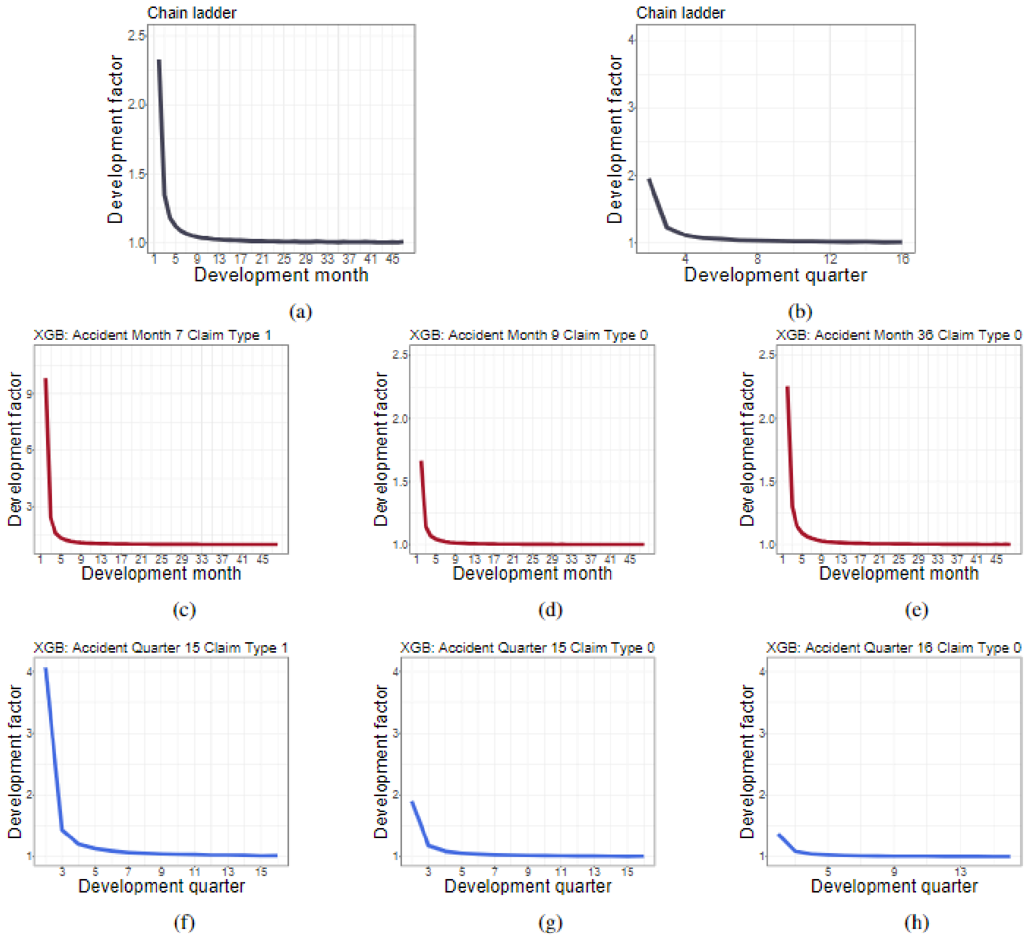
Article

- Hiabu, M., Hofman, E., & Pittarello, G. (2023). A machine learning approach based on survival analysis for IBNR frequencies in non-life reserving. arXiv:2312.14549.
- Hofman, E, Pittarello, G, Hiabu, M (2023). ReSurv: Machine learning based models for predicting IBNR frequency. R package version 0.0.2, https://github.com/edhofman/ReSurv.

Motivation

- Improve existing literature by introducing general framework for large set of features.
  - An individual model for reserving utilizing the theory in Hiabu (2017) and Pittarello, Hiabu, and Villegas (2023).
- Expanding the toolbox for reserving actuaries.
- Maintain interpretability through development factors.

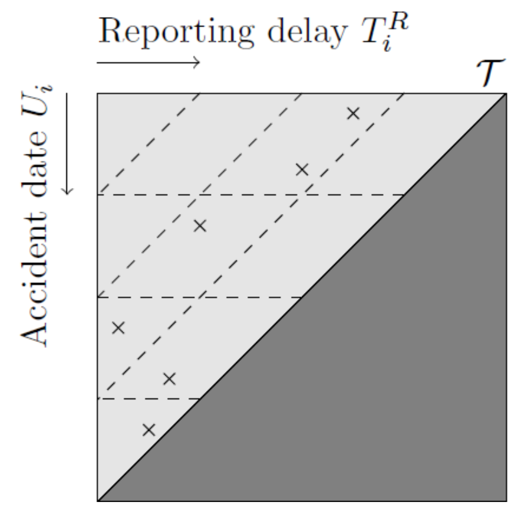
Result



Feature dependent development factors on monthly and quarterly granularity.

Modelling

We model claim reportings and build upon the theory in Hiabu (2017) and Pittarello, Hiabu, and Villegas (2023) using **survival models** for the **time-to-event analysis** of left-truncated and right-censored **non-life reserving** data.



Development triangle

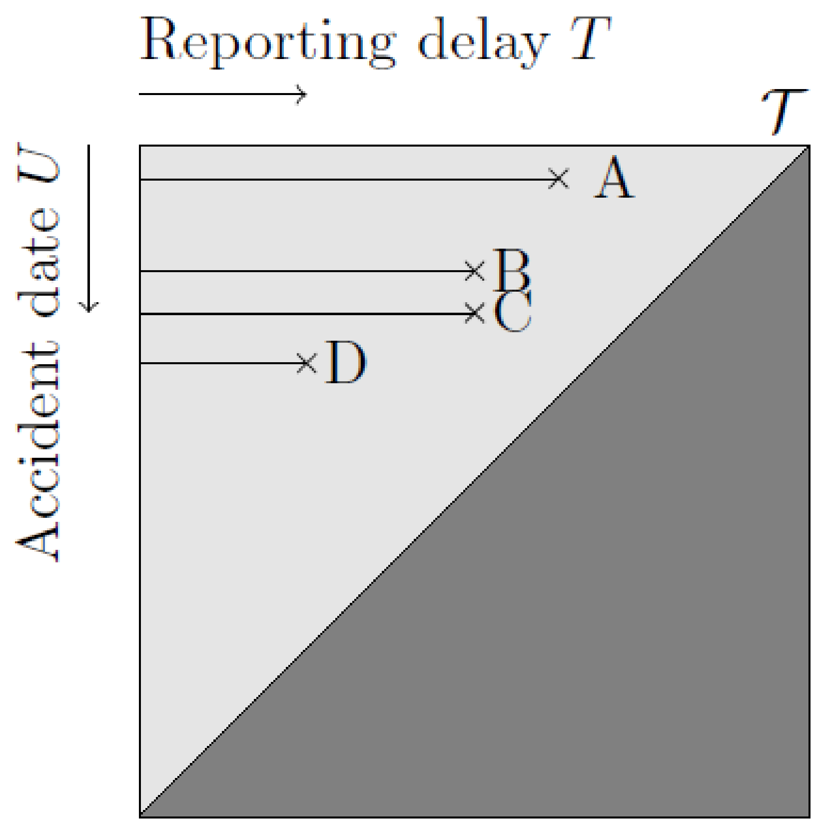
The i.i.d. data $(U_i, T_i^R), i = 1, \ldots, n$:

- U_i accident date.
- T_i^R the delay between accident and report.
- $\mathcal{T}$ is the evaluation date.

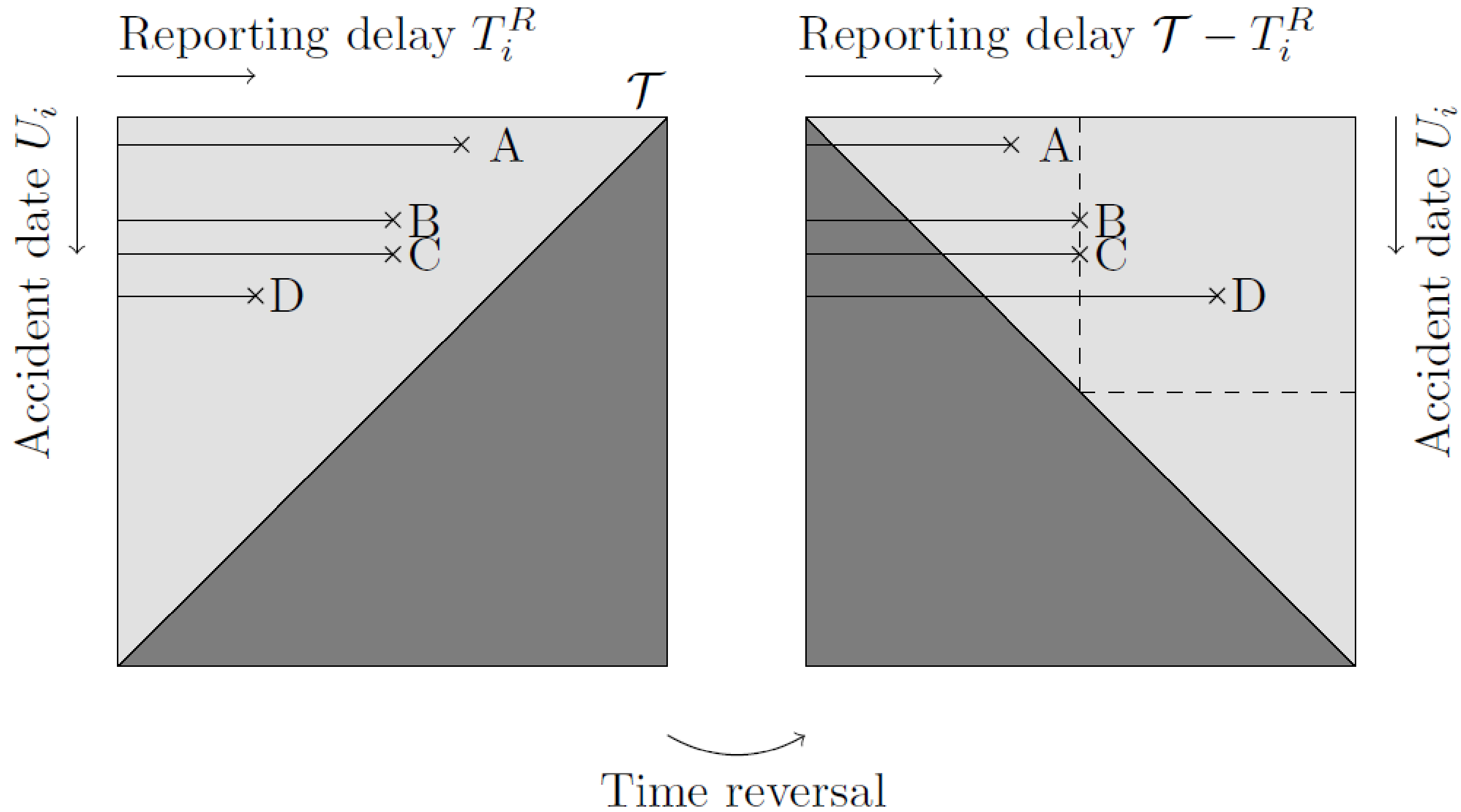
Reserving as a survival problem

Direct inference on T_i^R is not feasible, as T_i^R are observed up to the cut-off date $\mathcal{T}$. By design it holds

$$T_i^R \leq \mathcal{T} - U_i.$$



Four individual claims (A, B, C, D).



By not following every policy, we are exposed to a right-truncation problem. A solution to the right-truncation problem is to reverse the time leading to a tractable left-truncation problem, Hiabu (2017).

Under *usual conditions* (p.60, Andersen et al. (2012)), we consider the intensity of the development-time reversed counting processes $N_i'(t) = I(t \geq \mathcal{T} - T_i^R)$,

$$\lambda_i'(\mathcal{T} - t|U_i, X_i) = \lim_{h \downarrow 0} h^{-1} E\left[N_i'\{(\mathcal{T} - t + h)-\} - N_i'(\mathcal{T} - t) - \mid \mathcal{F}_{i,(\mathcal{T}-t)-}\right]$$

Let

$$\lambda_i'(\mathcal{T} - t|U_i, X_i) = \alpha^R(t|U_i, X_i) Y_i(t),$$

with $Y_i(t) = I(T_i^R \leq t < \mathcal{T} - U_i)$.

We model the hazard function

$$\alpha^{R}(t|u,x)=\lim_{h\downarrow 0} h^{-1}P\left(T_i^R\in(t-h,t]|\, Y_i(t)=1, X_i=x_i, U_i=u\right).$$

We propose to model

$$\alpha^{R}(t|u,x)=\alpha_0^R(t)e^{\phi(x,u;\theta)},$$

where $\alpha_0^R(t)$ is called the **baseline hazard** and $e^{\phi(x,u;\theta)}$ is the **risk score**; a component that depends on the features X_i and the accident period U_i and some parameters θ.

Using a proportional model, we minimize the partial-likelihood at time $t^{(\ell)}$

$$\mathcal{L}(X_1,\dots,X_n;U_1,\dots,U_n;\theta)=$$

$$\prod_{\ell=1}^{m}\prod_{i\in\mathcal{O}(t^{(\ell)})}\frac{e^{\phi(X_i,U_i;\theta)}}{\sum_{k\in\mathcal{R}(t^{(\ell)})}e^{\phi(X_k,U_k;\theta)}-\frac{\psi_i(\ell)}{O_\ell}\sum_{s\in\mathcal{O}(t^{(\ell)})}e^{\phi(X_s,U_s;\theta)}},$$

to obtain an estimate $\hat{\theta}$ of θ.

We have the exposure set

$$\mathcal{R}(t^{(\ell)})=\left\{i\in\{1,\dots,n\}:\mathcal{T}-t_i<t^{(\ell)}\cap\mathcal{T}-t_i\geq t^{(\ell)}\right\}$$

and the occurrence set

$$\mathcal{O}(t^{(\ell)})=\left\{i\in\{1,\dots,n\}:t^{(\ell)}\leq\mathcal{T}-t_i<t^{(\ell+1)}\right\},$$

while we indicate with $O_\ell=\#\mathcal{O}(t^{(\ell)})$ the cardinality of the set $\mathcal{O}(t^{(m)})$.

Finally $\psi_i(\ell)=i-\sum_{q=1}^{\ell-1}O_q-1$.

Using a proportional model, we minimise the partial-likelihood at time $t^{(\ell)}$

$$\mathcal{L}(X_1, \dots, X_n; U_1, \dots, U_n; \theta) = \prod_{\ell=1}^{m} \prod_{i \in \mathcal{O}(t^{(\ell)})} \frac{e^{\phi(X_i, U_i; \theta)}}{\sum_{k \in \mathcal{R}(t^{(\ell)})} e^{\phi(X_k, U_k; \theta)} - \frac{\psi_i(\ell)}{O_\ell} \sum_{s \in \mathcal{O}(t^{(\ell)})} e^{\phi(X_s, U_s; \theta)}}.$$

We extended the current survival analysis algorithms to model left-truncated data.

Model	**Effects modeling**	**Reference**
COX	$\phi\left(x, u; \theta\right) = \sum_{l=1}^{p} \theta_l X_l + \zeta\left(U\right)$	Gray (1992)
NN	$\phi\left(x, u; \theta^{NN}\right)$	Goodfellow, Bengio, and Courville (2016)
XGB	$\phi\left(x, u; \theta\right) = f_0(u_i) + f_1(x_i) + \ldots + f_K(x_i)$	Chen and Guestrin (2016)

Estimation of $\phi\left(x, u; \theta\right)$

We extended the current survival analysis algorithms to model left-truncated data.

Model	Effects modeling	Reference
COX	$\phi\left(x, u; \theta\right) = \sum_{l=1}^{p} \theta_l X_l + \zeta\left(U\right)$	Gray (1992)
NN	$\phi\left(x, u; \theta^{NN}\right)$	Goodfellow, Bengio, and Courville (2016)
XGB	$\phi\left(x, u; \theta\right) = f_0(u_i) + f_1(x_i) + \ldots + f_K(x_i)$	Chen and Guestrin (2016)

Estimation of $\alpha_0^R(t)$

Assuming uniform occurrences in the ties, we propose the following estimator for the baseline

$$\hat{\alpha}^R_{0,t^{(\ell)}} = \frac{O_\ell}{\sum_{k \in \mathcal{R}(t^{(\ell)})} e^{\phi\left(x_k, u_k; \hat{\theta}\right)} - 0.5 \sum_{s \in \mathcal{O}(t^{(\ell)})} e^{\phi\left(x_s, u_n; \hat{\theta}\right)}}.$$

Modelling the development factors

For $j, k = 1, \dots, m$, we propose to use the relation in Pittarello, Hiabu, and Villegas (2023) to estimate the development factors as

$$\hat{f}_{k,j}(x) = \frac{2 + \hat{\alpha}^R(j|k,x)}{2 - \hat{\alpha}^R(j|k,x)}.$$

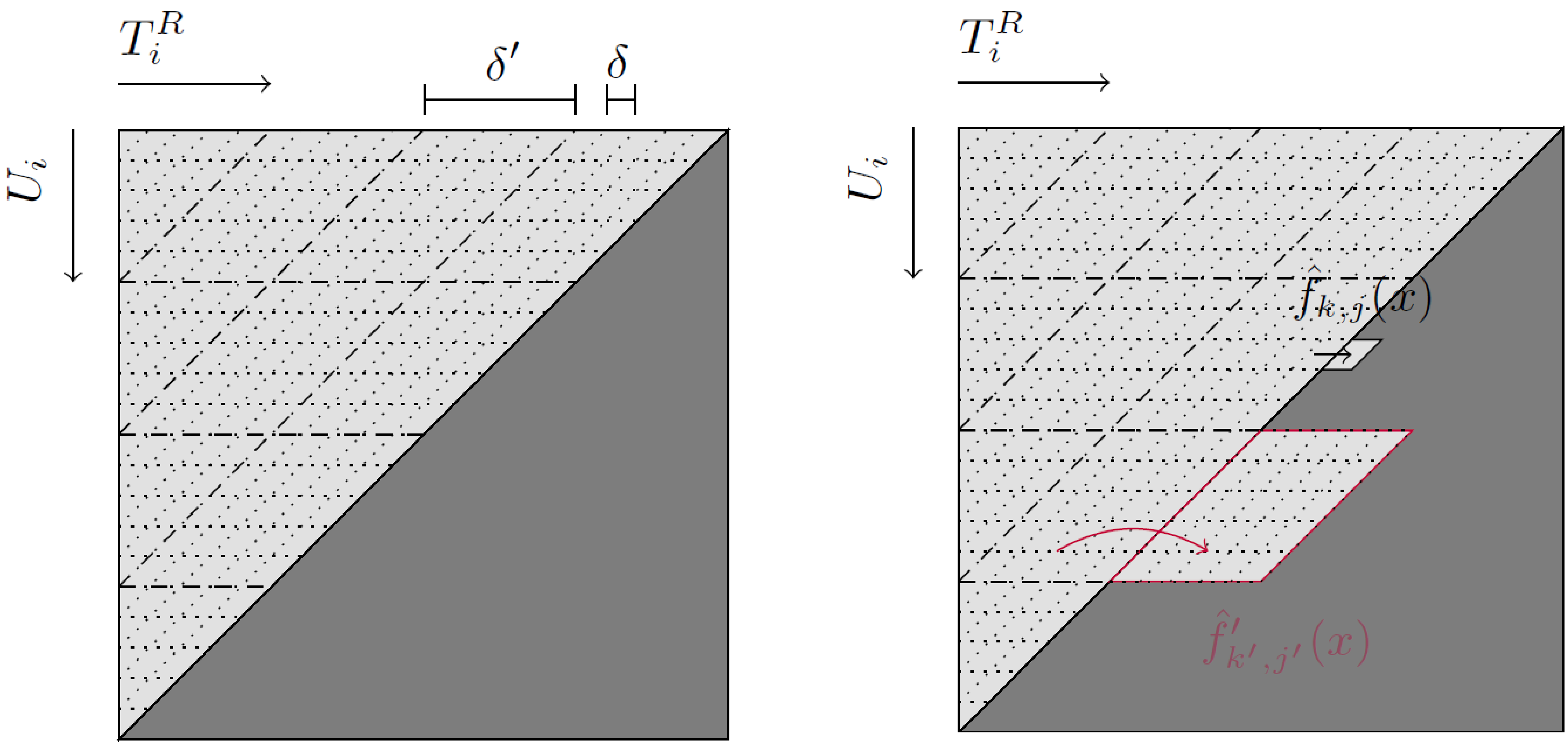
For $j, k = 1, \ldots, m$, we propose to use the relation in Pittarello, Hiabu, and Villegas (2023) to estimate the development factors as

$$\hat{f}_{k,j}(x) = \frac{2 + \hat{\alpha}^R(j|k,x)}{2 - \hat{\alpha}^R(j|k,x)}.$$

In Hiabu, Hofman, and Pittarello (2023), we show that we can obtain $\hat{f}'_{k,j}(x)$ for a granularity δ', with $\delta' > \delta$ and $\lfloor \frac{\delta'}{\delta} \rfloor = 0$.

For $j, k = 1, \ldots, m$, we propose to use the relation in Pittarello, Hiabu, and Villegas (2023) to estimate the development factors as

$$\hat{f}_{k,j}(x) = \frac{2 + \hat{\alpha}^R(j|k,x)}{2 - \hat{\alpha}^R(j|k,x)}.$$



Simulated data

We evaluate our models on five simulated scenarios, with same data composition.

Covariates	Description
Claim_number	Policy identifier.
Claim_type	Type of claim.
AD	Accident day.
RD	Reporting day.
DD	Development day.

There are different effects on ϕ

Scenario	Effect(s) on ϕ	CL	COX	NN	XGB
Alpha	claim_type	✓	✓	✓	✓
Beta	claim_type	✗	✓	✓	✓
Gamma	claim_type + claim_type: $\sqrt{AD}$	✗	✗	✓	✓
Delta	claim_type + AD	✗	✓	✓	✓
Epsilon	claim_type	✓	✗	✗	✗

There are different effects on ϕ

Scenario	Effect(s) on ϕ	CL	COX	NN	XGB
Delta	claim_type + AD	✗	✓	✓	✓
Epsilon	claim_type	✓	✗	✗	✗

Scenario delta includes seasonality in claim reporting. Epsilon breaks the model assumptions.

We evaluate our models on five simulated scenarios, with same data composition.

Covariates	Description
Claim_number	Policy identifier.
Claim_type	Type of claim.
AD	Accident month.
RD	Reporting month.
DD	Development month

Modelling strategy

For each scenario (Delta, Epsilon), we show the **average** results over 20 simulations.

For each simulation:

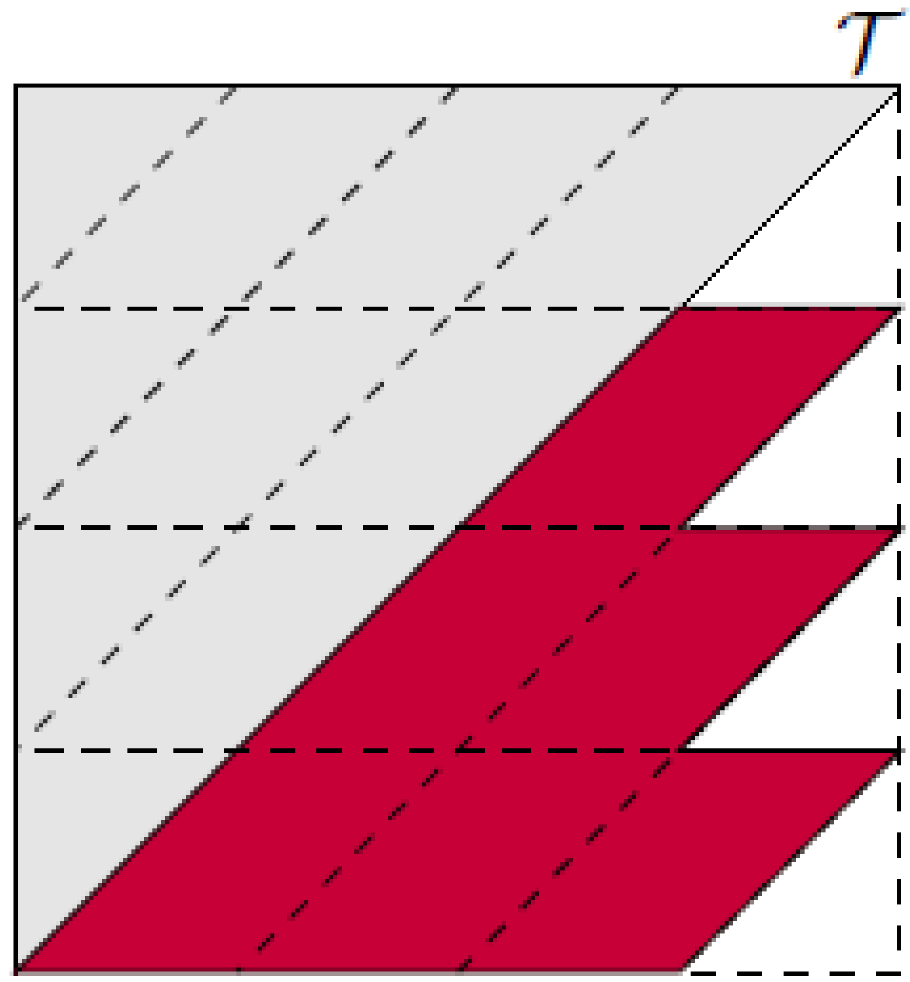
1. We optimise COX, NN, XGB hyper-parameters using a Bayesian approach (Snoek, Larochelle, and Adams (2012)).
2. We fit our models.
3. We predict future IBNR.
4. We evaluate the models performances.

- The (average) Continuously Ranked Probability Score (CRPS, Gneiting and Raftery (2007)).

$$\mathrm{CRPS}(\hat{S}(z|X,U),y) = \int_0^\infty (\hat{S}(z|X,U) - I\{y > z\})^2 \mathrm{d}z,$$

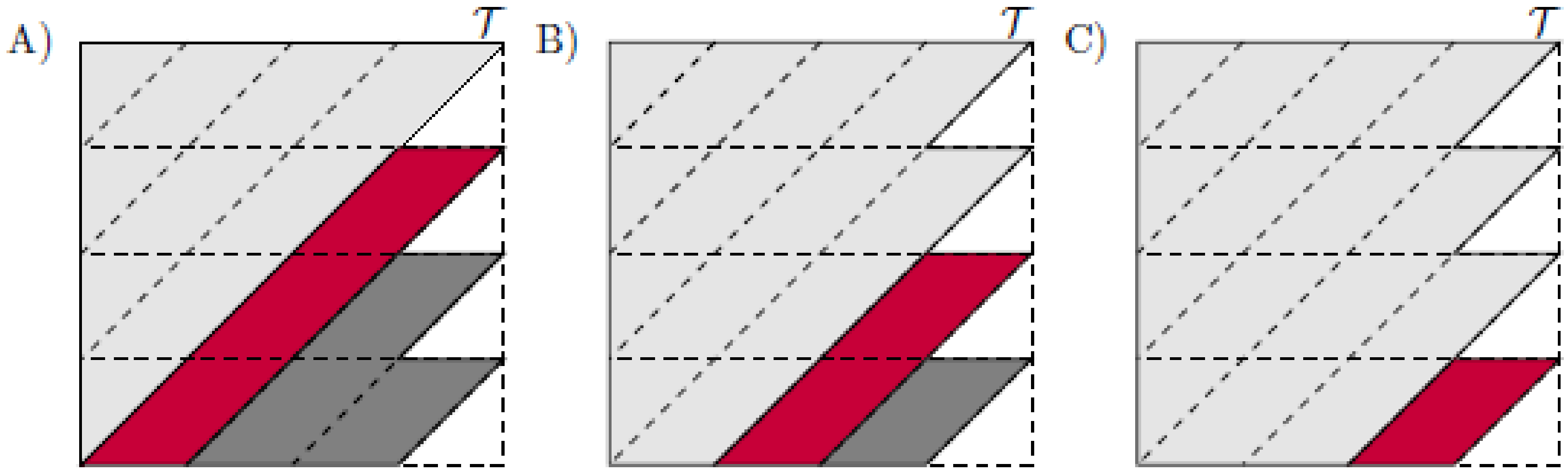
$$\hat{S}(j|X,U) = \frac{1}{\prod_{l=1}^{j} \hat{f}_{k,l}(x)}.$$

- Absolute Total Error on future calendar years ($\mathtt{ARE}^{\mathtt{TOT}}$).



- The (average) Continuously Ranked Probability Score (CRPS, Gneiting and Raftery (2007)).

- Absolute Total Error on future calendar years, with updating information ($\mathtt{ARE}^{\mathtt{CAL}}$).



- Absolute Total Error on future calendar years ($\mathtt{ARE}^{\mathtt{TOT}}$).
- The (average) Continuously Ranked Probability Score (CRPS, Gneiting and Raftery (2007)).

$\mathtt{ARE}^{\mathtt{TOT}}$, and CRPS in scenario Delta

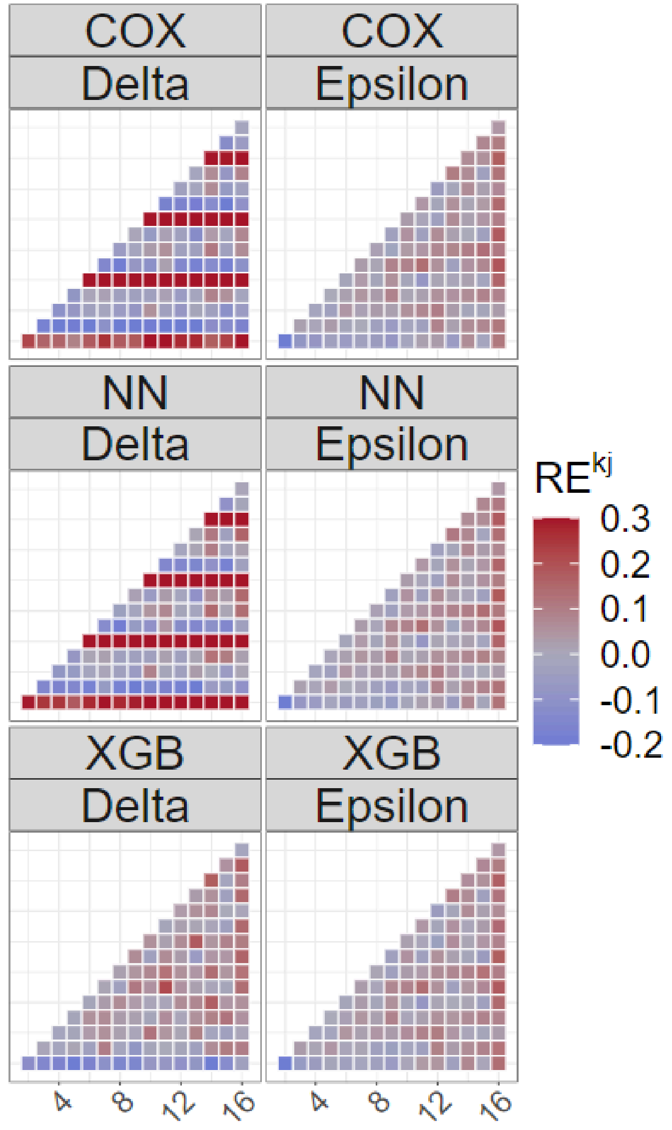
Model	$\mathtt{ARE}^{\mathtt{TOT}}$	CRPS
CL (✗)	0.300 ($\pm$ 0.024)	-
COX(✓)	0.195 ($\pm$ 0.022)	386.598 ($\pm$ 6.667)
NN (✓)	0.213 ($\pm$ 0.034)	388.707 ($\pm$ 7.205)
XGB (✓)	0.167 ($\pm$ 0.019)	369.531 ($\pm$ 6.599)

$\mathtt{ARE}^{\mathtt{TOT}}$, and CRPS in scenario Delta

Model	$\mathtt{ARE}^{\mathtt{TOT}}$	**CRPS**
CL (✗)	0.300 (± 0.024)	-
COX(✓)	0.195 (± 0.022)	386.598 (± 6.667)
NN (✓)	0.213 (± 0.034)	388.707 (± 7.205)
XGB (✓)	0.167 (± 0.019)	369.531 (± 6.599)

$\mathtt{ARE}^{\mathtt{CAL}}$ (quarters), and $\mathtt{ARE}^{\mathtt{CAL}}$ (years) in scenario Delta

Model	$\mathtt{ARE}^{\mathtt{CAL}}$ **(quarters)**	$\mathtt{ARE}^{\mathtt{CAL}}$ **(years)**
CL (✗)	0.234 (± 0.018)	0.037 (± 0.012)
COX(✓)	0.192 (± 0.018)	0.065 (± 0.020)
NN (✓)	0.204 (± 0.025)	0.051 (± 0.023)
XGB (✓)	0.145 (± 0.011)	0.058 (± 0.025)



The average (over the simulations) of the relative errors: $\frac{\sum_x O_{k,j}(x) - \sum_x \hat{O}_{k,j}(x)}{\sum_x O_{k,j}(x)}$

$\mathtt{ARE}^{\mathtt{TOT}}$, and CRPS in scenario Epsilon.

Model	$\mathtt{ARE}^{\mathtt{TOT}}$	**CRPS**
CL (✓)	0.119 ($\pm$ 0.011)	-
COX (✗)	0.135 ($\pm$ 0.022)	340.267 ($\pm$ 5.192)
NN (✗)	0.132 ($\pm$ 0.015)	341.169 ($\pm$ 5.210)
XGB (✗)	0.149 ($\pm$ 0.059)	340.344 ($\pm$ 5.100)

$\mathtt{ARE}^{\mathtt{CAL}}$ (quarters), and $\mathtt{ARE}^{\mathtt{CAL}}$ (years) in scenario Epsilon

Model	$\mathtt{ARE}^{\mathtt{CAL}}$ **(quarters)**	$\mathtt{ARE}^{\mathtt{CAL}}$ **(years)**
CL (✓)	0.115 ($\pm$ 0.010)	0.035 ($\pm$ 0.010)
COX (✗)	0.127 ($\pm$ 0.010)	0.060 ($\pm$ 0.015)
NN (✗)	0.126 ($\pm$ 0.011)	0.059 ($\pm$ 0.019)
XGB (✗)	0.127 ($\pm$ 0.012)	0.057 ($\pm$ 0.017)

Final remarks

- Introduced feature dependent development factors.
- Simulation study indicates methodology seems to work.
- Next step would be to include outstanding claim amounts.
- For case study and further simulations please refer to original manuscript and R-package.

Thank you for your attention!

Performance measures

- Absolute Total Error on future calendar years, with updating information ($\texttt{ARE}^{\texttt{CAL}}$) .

$$\texttt{ARE}^{\texttt{CAL}} = \frac{\sum_{\tau=M'+1}^{2M'-1} \sum_{j,k:k+j=\tau} \left| \sum_x O_{k,j}(x) - \sum_x \tilde{f}_{k,j}(x) O_{k,j-1}(x) \right|}{\sum_{j,k:k+j>M'} \sum_x O_{k,j}(x)}$$

- Absolute Total Error on future calendar years ($\texttt{ARE}^{\texttt{TOT}}$).

$$\texttt{ARE}^{\texttt{TOT}} = \frac{\sum_{j,k:k+j>M'} \left| \sum_x O_{k,j}(x) - \sum_x \hat{O}_{k,j}(x) \right|}{\sum_{j,k:k+j>M'} \sum_x O_{k,j}(x)}.$$

- The (average) Continuously Ranked Probability Score (CRPS, Gneiting and Raftery (2007)).

$$\mathrm{CRPS}(\hat{S}(z|X,U), y) = \int_0^\infty (\hat{S}(z|X,U) - I\{y > z\})^2 \mathrm{d}z,$$

$$\hat{S}(j|X,U) = \frac{1}{\prod_{l=1}^{j} \hat{f}_{k,l}(x)}.$$

Supplementary material A: Usual conditions

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Consider the filtration:

$$(\mathcal{F}_t : t \in \mathcal{T}) \, .$$

When the complete set of assumptions hold, we say that $\mathcal{F}_t$ satisfies the usual conditions (les conditions habitue lies):

- $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ for all $s < t$. (increasing)
- $\mathcal{F}_s = \bigcap_{t>s} \mathcal{F}_t$ for all s. (right-continuous)
- $A \subseteq B \in \mathcal{F}, \mathrm{P}(B) = 0 \Rightarrow A \in \mathcal{F}_0$. (complete)

We also require that the intensity $\lambda(\mathcal{T} - T_i^R)$ exists and is piecewise continuous.

References

Andersen, Per K, Ornulf Borgan, Richard D Gill, and Niels Keiding. 2012. *Statistical Models Based on Counting Processes*. Springer Science & Business Media.

Chen, Tianqi, and Carlos Guestrin. 2016. "Xgboost: A Scalable Tree Boosting System." In *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785–94.

Gneiting, Tilmann, and Adrian E Raftery. 2007. "Strictly Proper Scoring Rules, Prediction, and Estimation." *Journal of the American Statistical Association* 102 (477): 359–78.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press.

Gray, Robert J. 1992. "Flexible Methods for Analyzing Survival Data Using Splines, with Applications to Breast Cancer Prognosis." *Journal of the American Statistical Association* 87 (420): 942–51.

Hiabu, Munir. 2017. "On the Relationship Between Classical Chain Ladder and Granular Reserving." *Scandinavian Actuarial Journal* 2017: 708–29.

Hiabu, Munir, Emil Ditmar Hofman, and Gabriele Pittarello. 2023. "A Machine Learning Approach Based on Survival Analysis for IBNR Frequencies in Non-Life Reserving." *Preprint*.

Pittarello, Gabriele, Munir Hiabu, and Andrés M Villegas. 2023. "Chain Ladder Plus: A Versatile Approach for Claims Reserving." *arXiv Preprint arXiv:2301.03858*.

Snoek, Jasper, Hugo Larochelle, and Ryan P Adams. 2012. "Practical Bayesian Optimization of Machine Learning Algorithms." *Advances in Neural Information Processing Systems* 25.