Two Incompatible Objectives with Individual Reserve Models: an Approach with Multivariate Adaptive Regression Spline Models

Jean-Philippe Boucher

November 10, 2021

Co-operators Chair in Actuarial Risk Analysis
www.chairecara.uqam.ca
Universite du Quebec a Montreal (UQAM), Montreal
Canada

Introduction

Granular reserving approaches can be used in many ways:

1. Replace claims adjusters in setting individual case reserves for each open claim.
2. Tool when negotiating with insureds for a claim settlement;
3. Help claims managers to choose which action to take for each open claim;
4. Dynamically track the actuarial liabilities of any part of their portfolio;
5. Ratemaking teams could use the whole portfolio, without relying only on closed claims;
6. etc.

Objectives

We see that the examples do not share the same objectives.

A P&C insurance company is looking to satisfy the two following objectives:

- **O1** : The most accurate estimate of each individual case reserve at any time during the life of the claim.
- **O2** : An adequate value of the total reserve of the insurance portfolio, at any given time, to determine the company's actuarial liability.

Simple Example

Suppose a portfolio with only $n = 3$ claims, all reported at the same time:

- Claim #1, which will close at τ_1 and will cost \$200;
- Claim #2, which will close at τ_2 and will cost \$500;
- Claim #3, which will close at τ_3 and will cost \$2000;

with τ representing the age of the claim at closure, with $\tau_1 < \tau_2 < \tau_3$.

There are two main ways to assign individual case reserves for all open claims: a static and a dynamic estimation.

Static Estimation

General Idea: an individual reserve is assigned to each claim, and does not change over time.

- At time $x = 0$, each claim has a case reserve of \$900.
- Between time $[0, \tau_1[$, the sum of all case reserves is \$2700.
- At time $x = \tau_1$, claim # 1, reserved for \$900, is closed and paid at \$200.
- Between time $[\tau_1, \tau_2[$, the sum of all case reserves is now \$1800.
- At time $x = \tau_2$, claim # 2, reserved for \$900, is paid at \$500.
- Between time $[\tau_2, \tau_3[$, the sum of all case reserve is now \$900.
- At time $x = \tau_3$, claim # 3, reserved for \$900, is paid at \$2000.

Static Estimation (2)

O1 objective: the sum of all individual case reserves at their closure time (\$900 + \$900 + \$900) is equal to the total cost of claims (\$200 + \$500 + \$2000).

O2 objective: between τ_1 and τ_3, the sum of the individual case reserves is less than the actuarial liability.

Dynamic Estimation

General Idea: an individual reserve is assigned to each claim, but can change over time.

- At time $x = 0$, each claim is reserved for \$900.
- Between time $[0, \tau_1[$, the sum of all case reserves is \$2700.
- At time $x = \tau_1$, claim #1, reserved for \$900, is paid at \$200. **The remaining claims will cost \$2500: each open claim has a case reserve of \$1250.**
- Between time $[\tau_1, \tau_2[$, the sum of all case reserves is now \$2500.
- At time $x = \tau_2$, claim #2, reserved for \$1250, is paid at \$500. **The remaining claim will cost \$2000: the last claim is reserved at \$2000.**
- Between time $[\tau_2, \tau_3[$, the sum of all case reserves is now \$2000.
- At $x = \tau_3$, claim #3, reserved for \$2000, is paid at \$2000.

Dynamic Estimation (2)

O1 objective: the sum of all individual case reserves at their closure time (\$900 + \$1250 + \$2000) is higher than the total cost of claims (\$200 + \$500 + \$2000).

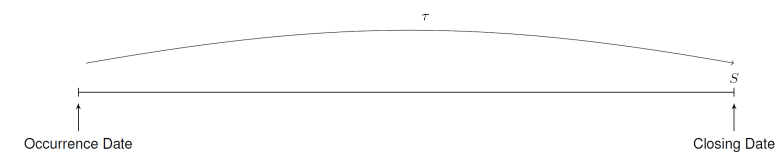
O2 objective: At all time, the sum of the individual case reserves was equal to the actuarial liability.

Plan of the Presentation

- Simulations to study two more sophisticated examples;
- We use a real insurance dataset to see how granular reserving models should be handled regarding the conflicting objectives;
- We develop a new individual reserve model based on Multivariate Adaptive Regression Spline (MARS) models, to model the relation between the age of the claim at closure and the value of the claim.
- We generalize the new MARS model to improve its precision.

Simulations

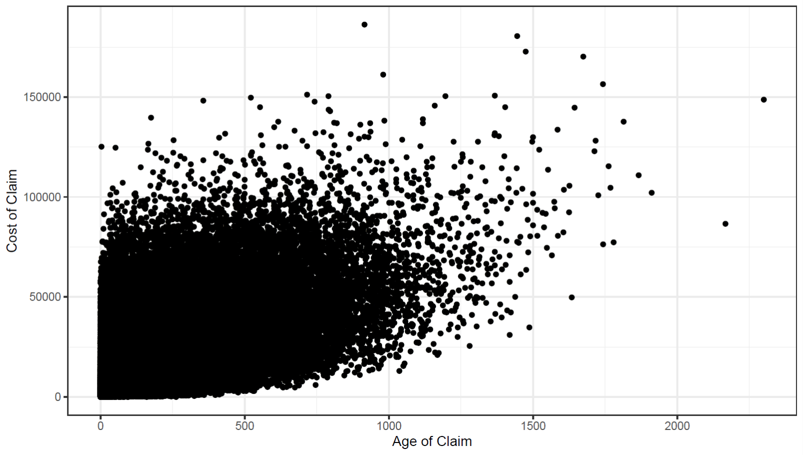
We suppose only one indemnity payment that occurs on the closing date of the claim:



Simulations (2)

- The age of the claim i at closure, τ_i, corresponds to the time between its date of occurrence and its date of closure.
- τ_i follows an exponential distribution of mean 180 days.
- The cost of a claim i, S_i, follows a gamma distribution of mean $\mu_i = 10,000 + 50\tau_i$, with shape parameter $\alpha = 10,000$.
- We simulate 10 claims per day, over a calendar time period ranging from 1 to $15,000$ days.

Results


150000
100000
50000
0
Cost of Claim
0
500
1000
1500
2000
Age of Claim

Calculation of Individual Reserves

1. Identical and unique reserves for all claims, at any age x (a.k.a. static reserve).

$$R_i^{(1)}(x) \quad = \quad \int_0^{\infty} (\beta_0 + \beta_1 t)\lambda \exp(-\lambda t) dt = \beta_0 + \frac{\beta_1}{\lambda}$$

2. Reserve a claim at any age x, as if x corresponds exactly to the time the claim is closed:

$$R_i^{(2)}(x) = \beta_0 + \beta_1 x$$

3. Using the conditional expectation of S_i to compute the reserve (a.k.a. dynamic reserve):

$$R_i^{(3)}(x) = E[S_i|\tau_i > x] = \int_x^{\infty} (\beta_0 + \beta_1 t) \frac{f(t)}{1 - F(x)} dt = \beta_0 + \beta_1 x + \frac{\beta_1}{\lambda}$$

Analyzing Each Model

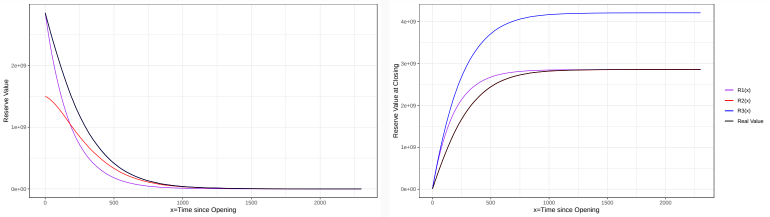


Figure 1: The graph on the left shows the total value of reserves for open claims, while the graph on the right shows the value of the reserve when the claim was closed

Analyzing Each Model (2)

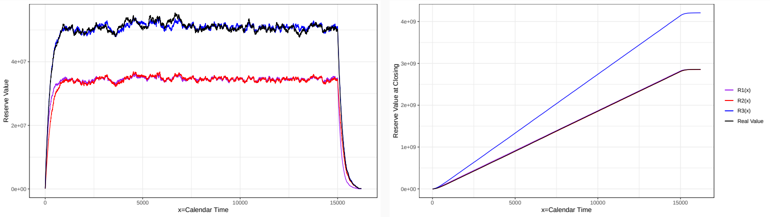


Figure 2: The graph on the left shows the total value of reserves for open claims, while the graph on the right shows the value of the reserve when the claim was closed

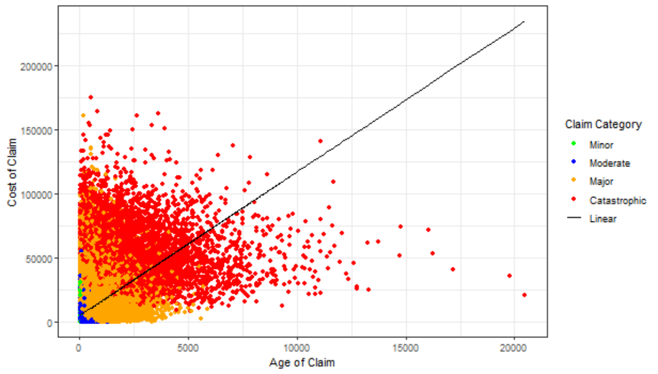
Simulations with Covariates

We have $\tau \sim \textit{Exponential}(\lambda)$ and $S|\tau \sim \textit{Gamma}(\mu = \beta_0, \alpha)$.

We introduce four types of claims:

j	Type of Claim	Proportion	λ	β_0	α
1	Minor	40%	$1/30$	$1,000$	$10,000$
2	Moderate	30%	$1/100$	$2,500$	$10,000$
3	Major	25%	$1/600$	$20,000$	$10,000$
4	Catastrophic	5%	$1/2000$	$50,000$	$10,000$

Simulations with Covariates (2)



If the covariate identifying the type of claim is not used in the modelling, the age of the claim at closure becomes significant, and will have to be used to compute each case reserve.

Calculation of Individual Reserves

- **Approach A1**: covariates are not considered.

$$R_i^{(A1)}(x) \quad = \quad \widehat{\gamma}_0 + \widehat{\gamma}_1 \left(x + \frac{1}{\lambda_i} \right)$$

- **Approach A2**: covariates are not considered and the distribution of τ_i is unknown. A gamma distribution is supposed and estimated:

$$R_i^{(A2)}(x) \quad = \quad \widehat{\gamma}_0 + \widehat{\gamma}_1 \frac{\widehat{\alpha}}{\widehat{\delta}} \frac{S_{\tau_i}(x; \widehat{\alpha}+1, \widehat{\delta})}{S_{\tau_i}(x; \widehat{\alpha}, \widehat{\delta})}$$

with $S_\tau(x)$ the cumulative function of τ.

- **Approach B**: We consider covariates for the modeling:

$$R_i^{(B)}(x) = \beta_0(i),$$

Analyzing Each Model

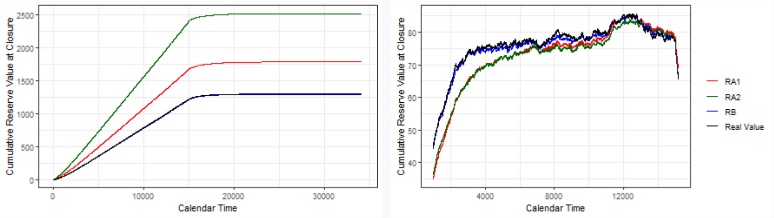


Figure 3: Cumulative reserve value at closure, for each calendar day (left), total reserve for each calendar day (right)

Statistics for Each Model

	Objective O1			Objective O2 (by day)			
	Diff (%)	MSE (Close)	MSE (Open)	Diff (%)	+/- 10%	+/- 5%	+/- 1%
$R^{(A1)}$	37.64%	11,886	11,360	-4.54%	84.24%	68.25%	11.80%
$R^{(A2)}$	94.15%	15,283	14,576	-5.14%	83.92%	62.47%	5.25%
$R^{(B)}$	-0.50%	9,303	9,303	-0.73%	100.00%	100.00%	55.59%

Impact of Incompatible Objectives

- A reserving model should include as many covariates as possible to minimize the use of the age of the claim in the modeling.
- When the age of the claim is still needed:
    - If not used in the model: we return to the static model $R^{(2)}$ which did not satisfy objective O2.
    - If we use it: we obtain dynamic model $R^{(3)}$ and objective O1 cannot be satisfied.

Impact of Incompatible Objectives: Practical Impacts

When a dynamic model is used:

1. If actuaries want to use the whole portfolio for ratemaking, we should expect bias.
2. The insurer must not judge the quality of their granular reserve models by the O1 criterion (contradicts what is currently asked of claims adjusters, who subjectively estimate all open claims).
3. It is important for the insurer not to rely solely on the value of the calculated case reserve if the insurer had to negotiate the settlement of a claim with one of its insureds.

Application with Real Insurance Data

1. We cannot know the impact of covariates on the cost of the claims, or the age of the claim at closure;
2. Some important variables might be missing;
3. The distribution of the age of the claim is not well defined;
4. The link between the cost of a claim and its age at closure can be caused by unobserved variables.

If the age of the claim has to be used as a covariate in the reserving model, it means that the best relationship between S and τ_i must be found.

Description of Data

We analyze the bodily-injury (BI) coverage, that has some particularities:

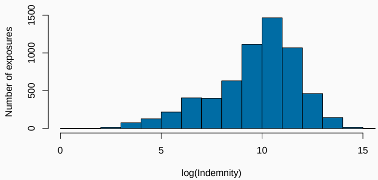
- A single accident can have multiple victims.
- A third-party (TP) is called an *exposure*, and each accident is called a *claim*. The micro-level reserving method must estimate the ultimate cost of each exposure.
- The insurance company do not have basic information about each TP.
- When the damages are paid to a TP, there is often a single and final payment of indemnity.

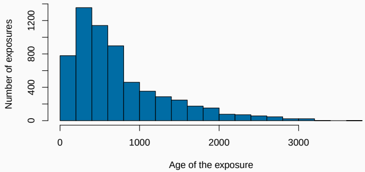
Working Assumptions

- We will only use claims with at least a single payment;
- We will suppose that all payments to an exposure are made at the closing date.
- We will not consider the potential dependence between exposures for the same claim.
- We will suppose that the covariates do not change over time, and are fully available at the time the claim is reported to the insurer.

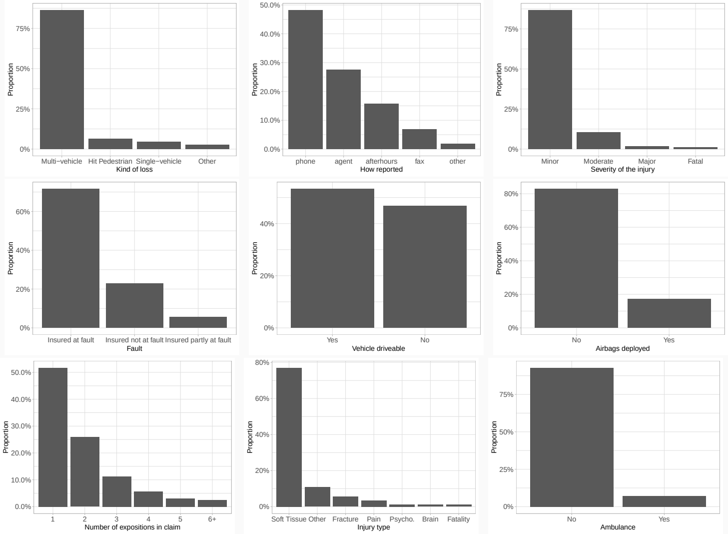
Basic Statistics for the indemnity and the age at closure:

	Average	Std.Err	Min.	Max.	25th pct.	50th pct.	75th pct.
Indemnity Paid	69,678	157,755	2	$\approx$ 4M	5,000	25,000	65,000
Age at closure	739.37	599.80	0	3,706	317	557	986





Available Covariates



Modeling Approach

For each exposure i, $i = 1, \ldots, n$, we are looking to use the following model:

$$R_i(x) \;=\; \frac{1}{1 - F_{\tau_i}(x)} \int_x^{\infty} E[S_i|\tau_i = t] f_{\tau_i}(t) dt$$

- We do not know the form of $E[S_i|\tau_i]$;
- We do not know the distribution of the random variable τ_i.

Our modeling strategy is thus a two-step approach: the age of the exposure at closure (τ_i) modeled first, and then the cost of exposure (S_i).

Constraints

To easily analyze the case reserves at any point of time, an analytical form of $R_i(x)$ would be better.

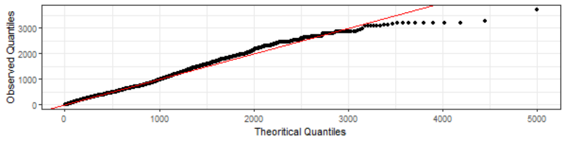
- At any time, a claim adjuster might be interested in the case reserve of a specific exposure;
- At any time, a claim manager may want to know the total liability value of the claim portfolio;
- etc.

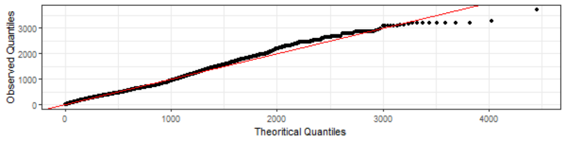
Models based on simulations, or on numerical approximations might not be a good solution as they cannot generate a number quickly enough.

To obtain an analytic form, we have to put some constraints on the distribution of τ_i as well as on the form of $E[S_i|\tau_i]$.

Age of Exposure at Closure (τ)

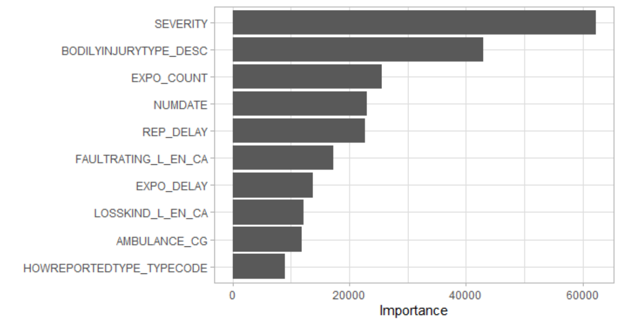
QQplots for two distributions: gamma and Weibull.





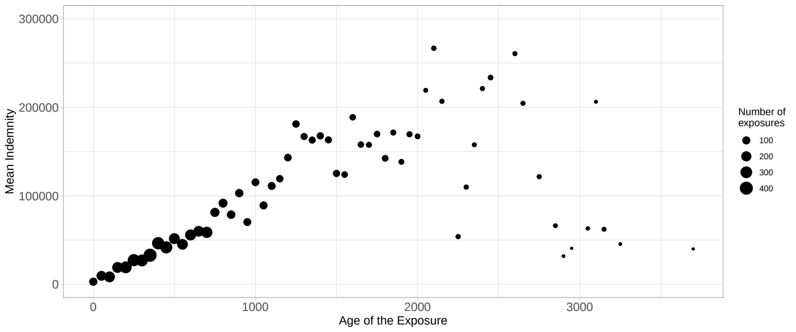
Covariates for τ

We chose to use a random forest approach, with hyper-parameters calibrated by cross-validation. The following figure shows the most important variables:



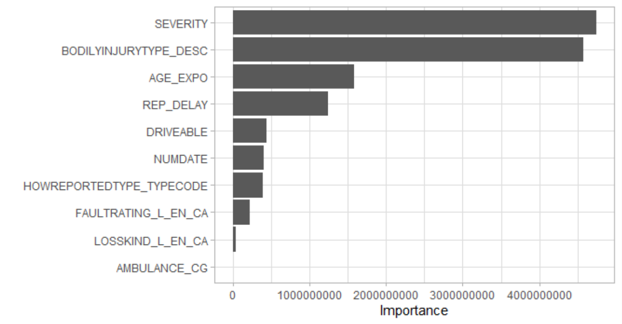
Cost of Exposure (S)

We first analyse the relationship between the cost of the exposure and the age at closure.



Covariates for S

With the default random forest model, the following figure shows the most important variables:



Parametric Form

A simple possibility is to use the following model:

$$E[S_i|\tau_i = t] = X_i'\delta + \gamma_1 t$$

where τ_i is the age of the exposure at closure. Covariates could be included in $X_i'\delta$ and γ_1 is the linear trend for the age at closure. It results in:

$$R_i(x) = \frac{1}{1 - F_{\tau_i}(x;\alpha,\beta)}(X'\delta S_{\tau_i}(x;\alpha,\beta) + \gamma_1 \frac{\alpha}{\beta} S_{\tau_i}(x;\alpha+1,\beta))$$

with $F_\tau(x), S_\tau(x)$, the cumulative and the survival functions respectively. With exponential, gamma and Weibull distributions, an analytical form of $\int_x^\infty t f_{\tau_i}(t)dt$ can easily be found.

Multivariate Adaptive Regression Splines (MARS)

To generalize, we could add a term t^2 to the estimation of the model, but instead we chose to use multivariate adaptive regression splines (MARS) models.

The MARS model is a non-parametric technique that supposes the following weighted sum for the mean parameter:

$$E[S] = \sum_{i=1}^{p} c_i B_i(y)$$

where each $c_i, i = 1, \ldots, p$ are parameters to be estimated.

Hinge Functions

The term $B_i(y)$ is a function of covariate y.

For categorical covariate y, $B_i(y)$ represents a dummy function (equals 1 or 0 depending on the value).

For numeric covariates, such as τ, $B_i(\tau)$ is a hinge function $h(.)$, with:

$$\begin{aligned} B_i(\tau) &= h(a_i - \tau) = \max(a_i - \tau, 0) \\ &\text{or} \\ B_i(\tau) &= h(\tau - a_i) = \max(\tau - a_i, 0) \end{aligned}$$

for a specific value of a_i estimated from the data.

Reserve Value

We use the following structure to model the cost of the exposure:

$$E[S_i|\tau] = \sum_{j=1}^{n_1} c_j^{(1)} B_j(y) + \sum_{j=1}^{n_2} c_j^{(2)} B_j(\tau).$$

With $B_k(t)$ a hinge function, we can compute $\int_x^\infty E[S(t)] f_\tau(t) dt$:

$$R(x) \quad = \quad \frac{1}{1 - F_\tau(x)} \left(\sum_{j=1}^{n_1} c_j^{(1)} B_j(x) S_\tau(x) + \sum_{j=1}^{n_2} c_j^{(2)} \int_x^\infty B_j(t) f_\tau(t) dt \right)$$

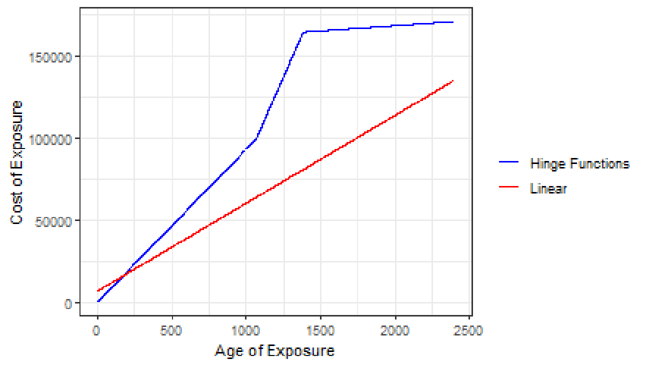
Reserve Value (2)

Each elements of the last equation can be solved easily. For example, we could have:

$$
\begin{aligned}
\int_x^{\infty} B_k(t) f(t) dt &= \int_x^{\infty} h(a_k - t) f(t) dt \\
&= \int_x^{\infty} \max(a_k - t, 0) f_\tau(t) dt \\
&= \begin{cases} \int_x^{a_k} (a_k - t) f_\tau(t) dt & \text{if } x < a_k \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

MARS model with Bodily Injury Data

To illustrate the way MARS can be used in our case, we directly apply a simple MARS model to the data, without any calibration.



Tuning the Model

We found all hyper-parameters of the reserving model in two steps.

We first found the hyper-parameters of the random forest model by cross-validation. For the MARS model, we separated the validation dataset into binds. For each estimation step:

1. For a specific couple (*degree/nprune*), we fitted the MARS model.
2. An automated integration algorithm, flexible enough to be used for all possible values of a for hinge functions was developed to compute each case reserve for all of the portfolio's exposures.
3. The reserve value of each exposure i, for any time x, can be computed.

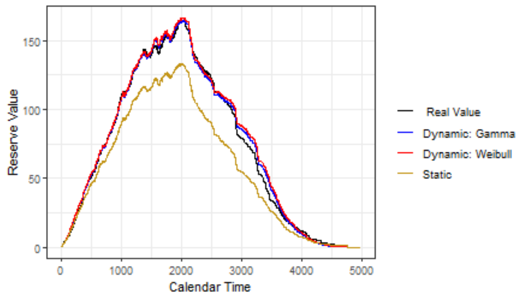
Hyper-parameters were found by selecting the best model.

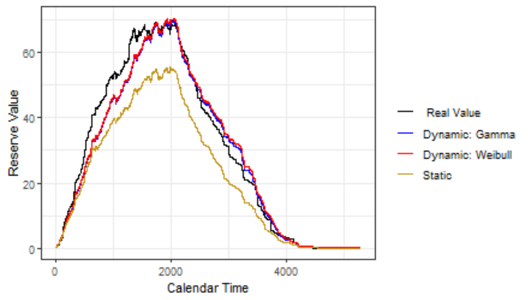
Analyzing the Reserving Model

A MARS model with 12 parameters without interaction applied to the cost of exposure generates the estimators shown in the following.

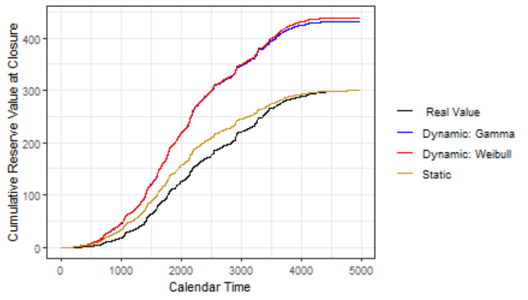
Parameter	Coefficient	Parameter	Coefficient
(Intercept)	538,598	-	.
Sev(Moderate)	-259,350	h(τ-2513)	45,643
Sev(Minor)	-293,620	h(τ-2543)	-62,479
Inj(Soft Tissue)	-88,297	h(τ-2631)	16,725
Inj(Pain)	-56,23	h(2543-τ)	-66
Inj(Other)	-57,407	-	.
Inj(Fracture)	-42,259	-	.
Inf(Fatality)	-203,169	-	.

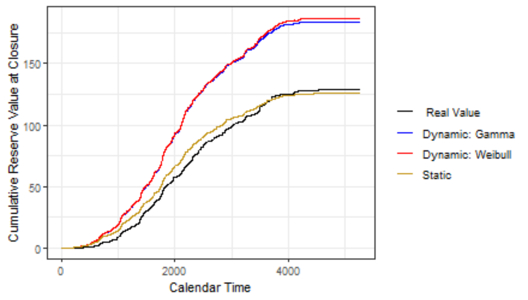
Results (training and test datasets)



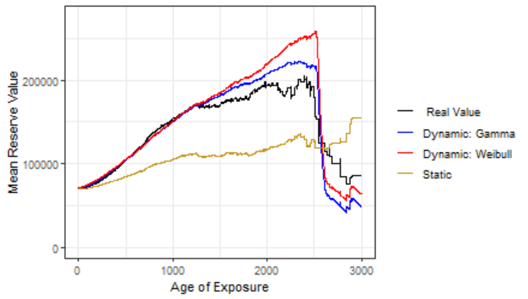


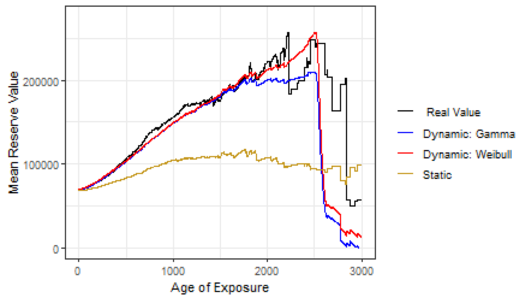
Results (training and test datasets)





Results (training and test datasets)





Importance of the Age of Exposure

Although the age of exposure at closure is an important variable in the modeling, it seems reasonable to believe that there is no real causal relationship between the time at closure and the final cost of the exposure.

If we observe an increasing trend, it is simply because it takes more time to close complex exposures that cost, on average, a lot more than other type of exposures.

The exposure at closure is a proxy for the complexity of the exposure. If the insurer could gather more precise information on each exposure, it will help identify complex exposures, which in turn will diminish the importance of the time to closure in the model.

Reducing the Effect of the Age of Exposure

We could first model $E[S_i|\tau]$ with only covariates that do not depend on the age of the exposure:

$$E[S_i] = \sum_{j=1}^{n_1} c_j^{(1)} B_j(y)$$

The approach is to fit a model on the residuals $W_i = S_i - \widehat{S}_i$ by including only covariates linked with the age of exposure at closure:

$$E[W_i|\tau] = \sum_{j=1}^{n_2} c_j^{(2)} B_j(\tau) + \sum_{j=1}^{n_3} c_j^{(3)} B_j(\tau, y).$$

where $B_j(\tau)$ is function of τ, and $B_j(\tau, y)$ is a function of τ and y.

Results

Original MARS models:

	Objective O1			Objective O2 (by day)			
	Diff (%)	MSEP (Close)	MSEP (Open)	Diff (%)	+/- 10%	+/- 5%	+/- 1%
Gamma	43.33%	163,077	160,615	11.58%	44.40%	25.51%	7.51%
Weibull	45.61%	162,948	160,624	12.94%	41.82%	24.34%	6.57%

MARS models on residuals:

	Objective O1			Objective O2 (by day)			
	Diff (%)	MSEP (Close)	MSEP (Open)	Diff (%)	+/- 10%	+/- 5%	+/- 1%
Gamma	25.32%	161,390	159,506	10.54%	54.79%	42.95%	11.55%
Weibull	28.66%	160,960	159,489	9.99%	57.65%	44.22%	9.38%

Conclusion

- When available covariates cannot explain the trend in the average cost, it is impossible to have a granular reserving model that provides [O1] an adequate prediction of the amounts paid for each claim, and [O2] an adequate value for the total actuarial liability.
- We need to develop a flexible approach to model the link between the age of the exposures at closure and the cost of the exposure.
- A new MARS model, tuned by a special procedure, was developed.
- A generalization of the MARS model was proposed, where the age of the exposure is only used on the residuals. This correction produces better results, and it comes closer to satisfying objectives O1 and O2.

References

- J.-P. Boucher (2021). Two Incompatible Objectives with Individual Reserve Models: An Approach with Multivariate Adaptative Regression Spline Models. *Submitted*.

- Co-operators Chair in Actuarial Risk Analysis: www.chairecara.uqam.ca