AKUR8

Penalized GLM : Between Credibility and GBMs

Mattia Casotto
Head of Product US

Biography

Mattia Casotto is the Head of Product for the United States division of the pricing software Akur8.

He has more than 7 years of experience on predictive modeling in insurance and is one of the founding members of Akur8.

He is one of the co-author of the white-paper ‘Credibility and Penalized Regression’.

AKUR8

Advantages of Penalized GLMs

Blending Credibility and incorporating non-linearities

Our aim is to introduce Penalized GLMs and highlight how they can address two major limitations of unpenalized GLMs

Blends Credibility with a GLM

- Penalized GLMs provide more robust estimates in segment with limited exposure

Natively fits non linear effects

- Robust nonlinearities are detected without feature engineering requirements

Part 1 - Penalized GLM and Credibility

How credibility leads to most robust estimates

This presentation is divided in **two parts**.

Blends Credibility with a GLM

The first section - **Credibility** - will illustrate:

- Why low exposure segments in a GLM may lead to non-sound estimates
- Current strategies to overcome this limitation via levels selection using significance analysis
- How credibility allows to leverage all available data
- How Penalized GLM can achieve both levels selection and credibility assumptions.

This results in several benefits

1. The resulting factor will be more sensible when there is a lack of data
2. The effort to question the significance of each factor is reduced since the factors are selected according to credibility

Part 2 - Penalized GLM and GBM

Natively fits non linear effects

The second section - GBM will illustrate

- An example of how non-linearities are modeled in GLM
- Why these techniques may lead to instabilities and volatile estimates
- How GBMs, a non-transparent modeling technique, incorporates non-linearities in the model via adaptive grouping
- How Penalized GLM can incorporate a similar adaptive grouping based on credibility

This results in several benefits

- Robust estimates of the risk, particularly on variables showing strong non-linearities
- Non-linearities are less prone to modeler choices and biases

Penalized GLMs in insurance

Where do they come from?

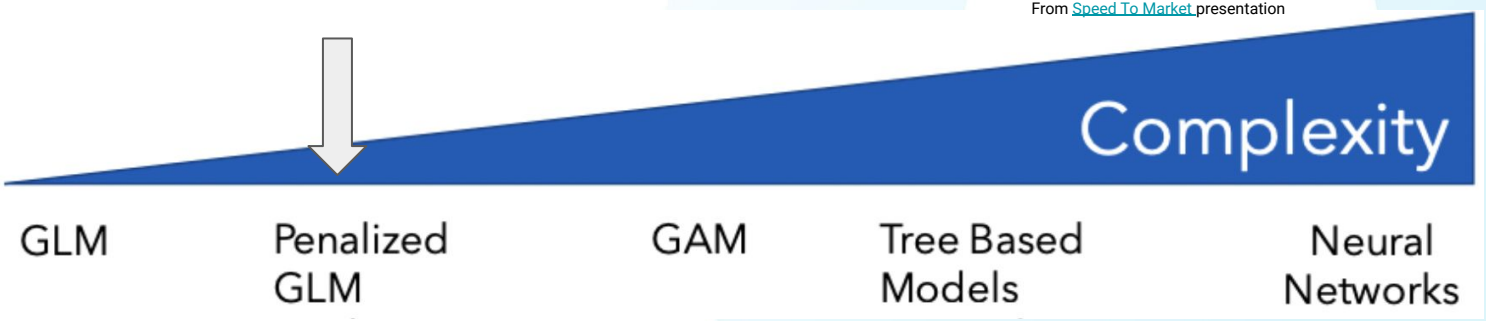
Penalized GLMs are a variation of the standard GLM, being studied and published for more than 20 years.

They are becoming increasingly popular in insurance applications:

Lasso, Ridge and Elastic net (Glmnet)

- Presented on the NAIC 2021 June Book club - Regularization Method.
- Presented on the NAIC 2022 October Book club - Pvalues and Alternatives
- Section 10.5 in the CAS Monograph - Generalized Linear Models for Insurance, Rating Second Edition
- Cited in Speed to Market presentation [April Book Club 2022] (below)

From Speed To Market presentation



Literature on Penalized GLM is growing

Several applications of Penalized regression in actuarial science.

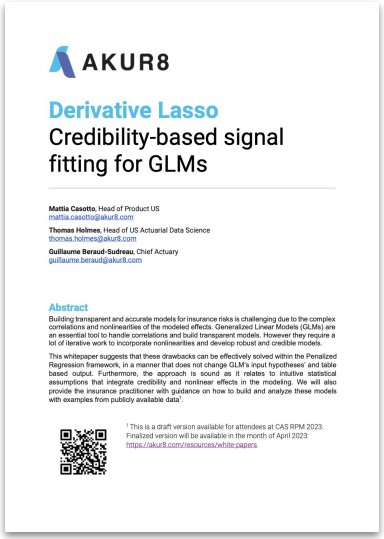
To inform and educate the market on how Penalized GLM work, on their similarities and difference from standard, unpenalized GLM, two main works are in progress:

Derivative Lasso: Credibility-based signal fitting for GLMs (2023 – Draft Available)

- The “Derivative Lasso” paper shows through a practical example with real data the difference (and limits) of the GLM methodology and how Penalized Regression address those limits so to enhance speed and accuracy of GLM models.

CAS Monograph: Penalized Regression as a Credibility Procedure (2023 - Draft in Progress)

- This monograph explores the mathematical connection between penalized regression and credibility. Additionally, it includes a practitioner’s guide to applying penalized regression as a credibility procedure in a loss model with accompanying code and guidance on model review.


AKUR8

Derivative Lasso

Credibility-based signal fitting for GLMs

Mattia Casotto, Head of Product US
mattia.casotto@akur8.com

Thomas Holmes, Head of US Actuarial Data Science
thomas.holmes@akur8.com

Guillaume Beraud-Sudreau, Chief Actuary
guillaume.beraud@akur8.com

Abstract

Building transparent and accurate models for insurance risks is challenging due to the complex correlations and nonlinearities of the modeled effects. Generalized Linear Models (GLMs) are an essential tool to handle correlations and build transparent models. However they require a lot of iterative work to incorporate nonlinearities and develop robust and credible models.

This whitepaper suggests that these drawbacks can be effectively solved within the Penalized Regression framework, in a manner that does not change GLM’s input hypotheses’ and table based output. Furthermore, the approach is sound as it relates to intuitive statistical assumptions that integrate credibility and nonlinear effects in the modeling. We will also provide the insurance practitioner with guidance on how to build and analyze these models with examples from publicly available data[1].

[1] This is a draft version available for attendees at CAS RPM 2023. Finalized version will be available in the month of April 2023. https://akur8.com/resources/white-papers

Credibility and Low Exposure Level

AKUR8

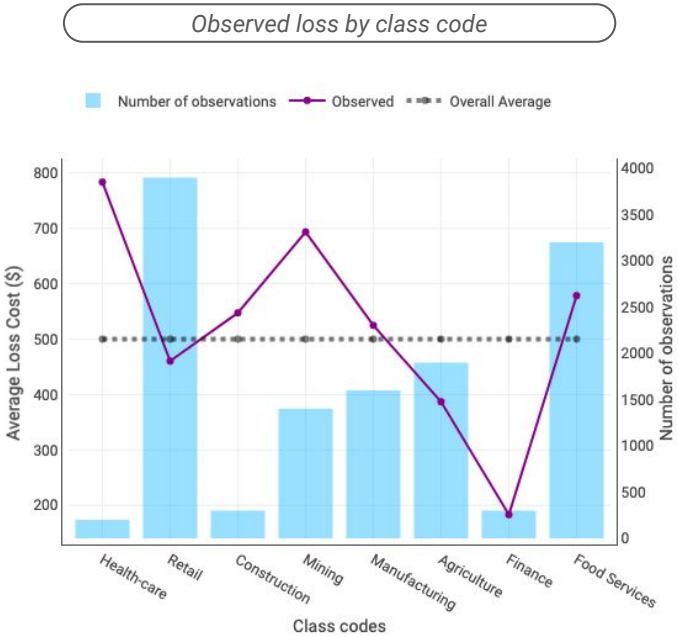
Worker's Compensation example

Loss Cost by class code example

Losses and exposures for companies are collected, and we want to compute an estimation of the average loss cost per class code.

The data **can be represented visually**:

- The **blue bars** represent the number of observations for a given class;
- The **purple lines** represent the **Observed Experience** as the average loss cost for each class;
- The **black line** represent the **overall average** (or grand average) of $500 in this example.



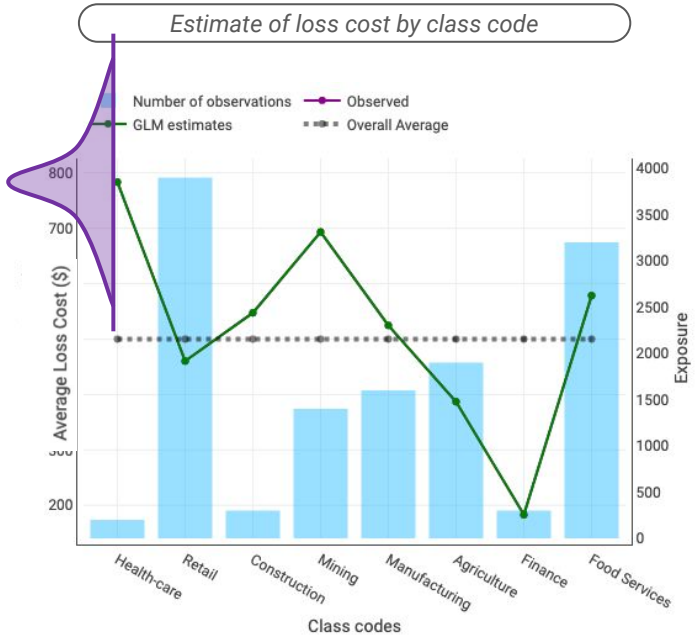
GLMs: Univariate estimate

The risk estimate can be computed in the GLM framework, by maximizing the Likelihood

$$\beta^* = Argmax\ Likelihood(Obs., \beta)$$

The resulting estimate is the average loss cost by class code.

However such estimate may be inappropriate for the class “Health-Care” which has low exposure.



GLMs: Univariate estimate

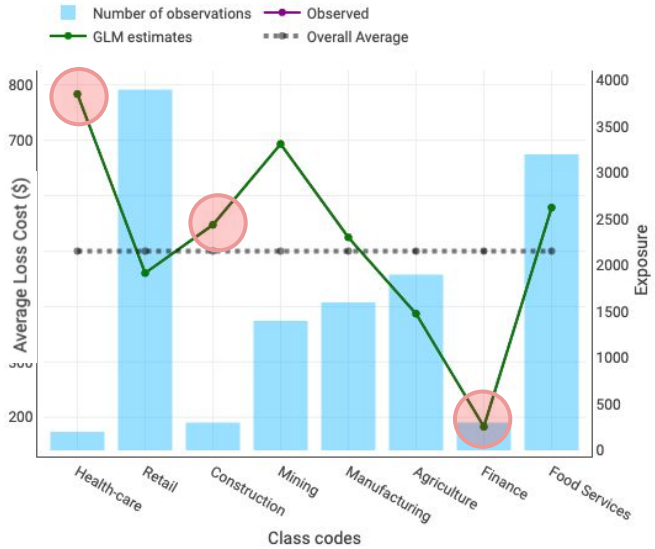
The risk estimate can be computed in the GLM framework, by maximizing the Likelihood

$$\beta^* = Argmax\ Likelihood(Obs., \beta)$$

The resulting estimate is the average loss cost by class code.

However such estimate may be inappropriate for the class “Health-Care” which has low exposure.

Estimate of loss cost by class code



Removing non-significant levels

AKUR8

Removing low-significance levels

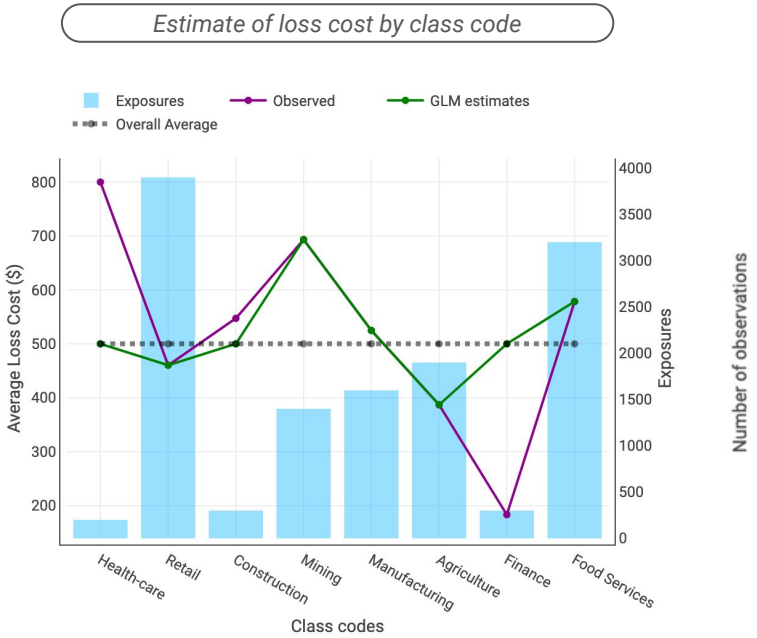
A classic approach is to use the **statistical significance** of the different levels, such as **p-values**.

Levels that have low exposure (or small effects) may be less significant, and their contribution may be set to zero (overall average).

The result obtained will depend on the **significance threshold** above which levels will be kept into the final model or grouped:

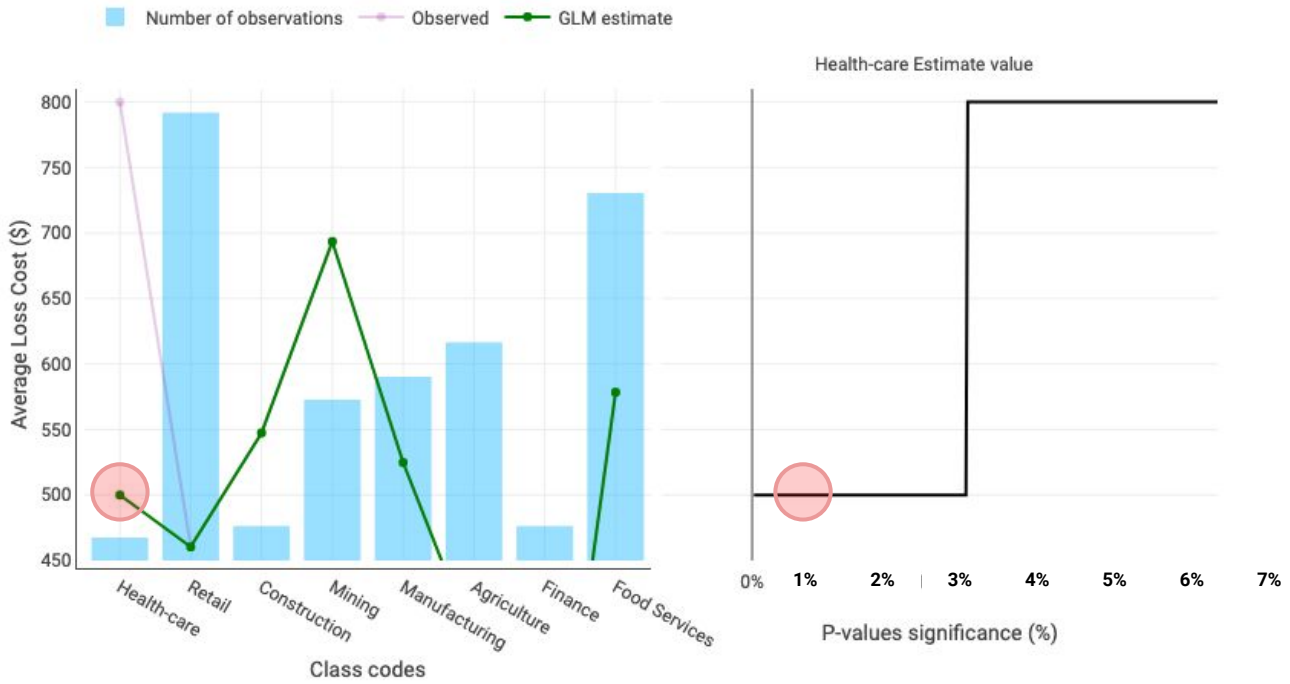
- If a level is **more significant** than the threshold, it is **kept;**
- If a level is **less significant** than the threshold, it is **removed.**

Modelers often use a “5% significance level” but other values can be selected.



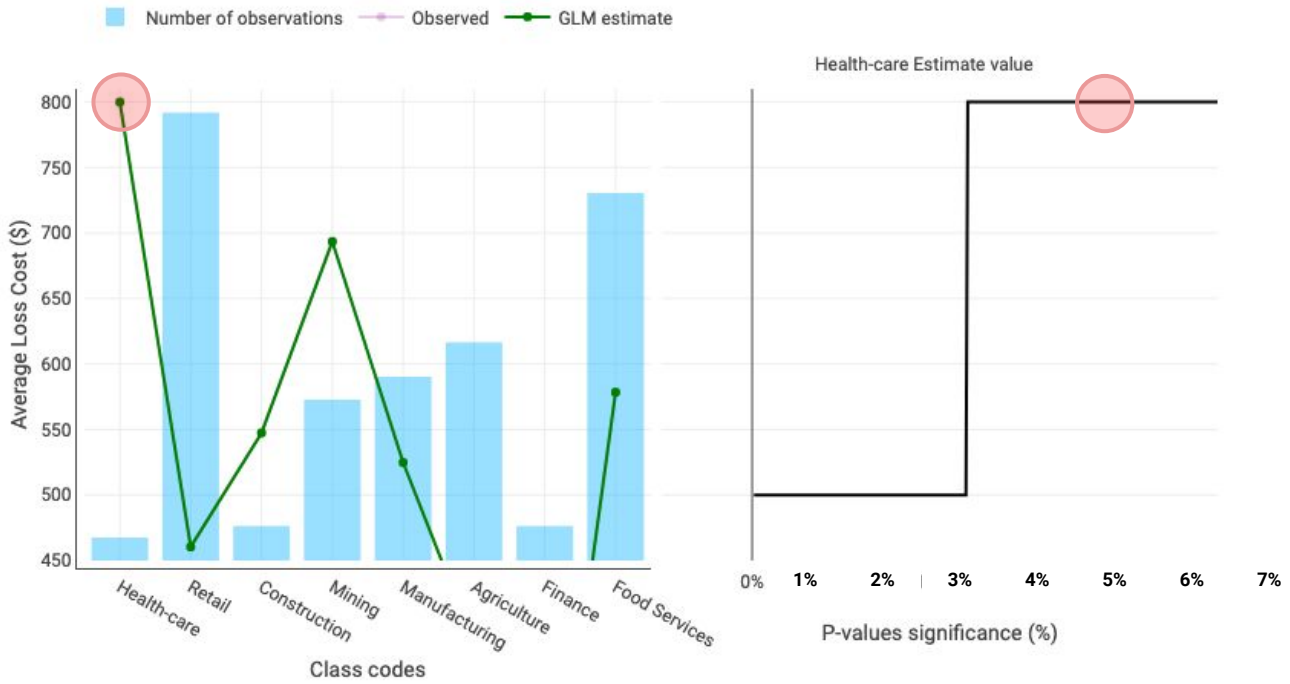
Fitted model depends on the threshold

Strong (low) significance thresholds are hard to validate and lead to a **robust** model that is less affected by noise in the data.



Fitted model depends on the threshold

Weak (high) significance threshold are easy to validate and lead to a **volatile** model that may be too reactive to noisy data.



Strengths & limits of levels selection

This approach has well known strengths and limits:

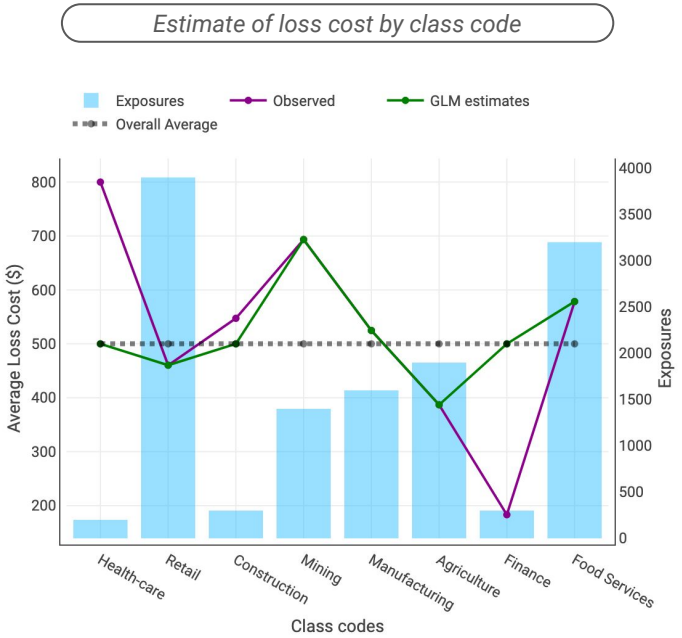
✅ It is a binary method, leading to clear decisions;

✅ It is very frequently used and widely accepted;

✅ It relies on very classic statistics.

❌ It is a binary method: it does not efficiently use the limited observations we have on “health-care”;

❌ The test’s justification relies on hypothesis that may not be met in practice.



Buhlmann Credibility

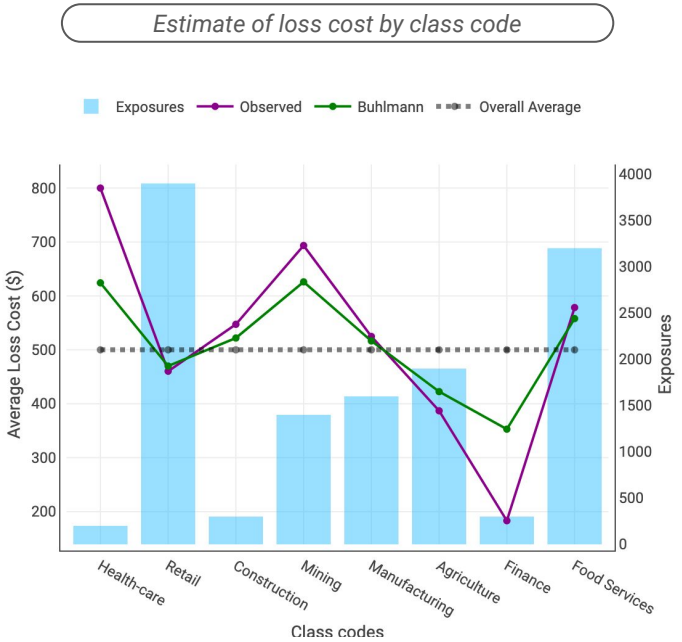
AKUR8

The Credibility solution

The idea of a credibility framework is to create non-binary predictions between these two extreme “yes” and “no” solutions.

Low-exposure levels are:

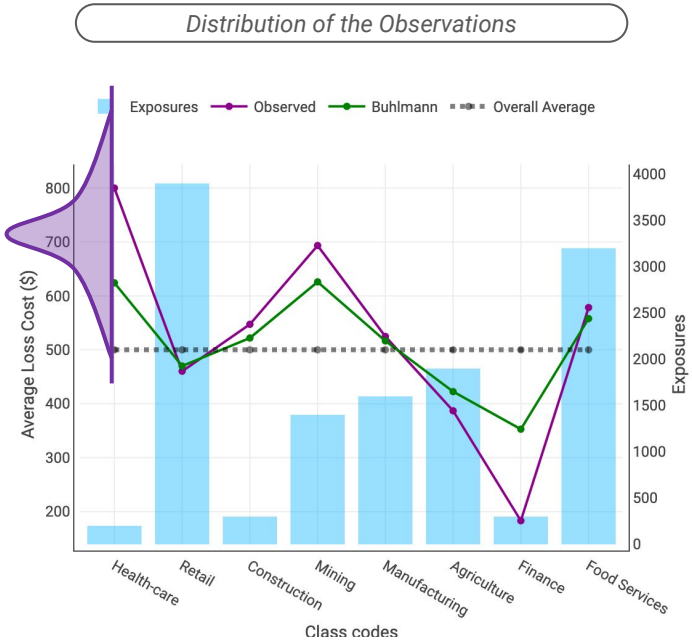
- **Not fully trusted** (like they would in a standard GLM framework);
- **Not fully discarded** (like they would if we applied a grouping of non-significant levels).



What is the idea motivating Credibility?

The Bühlmann credibility creates predictions by mixing two sources of information:

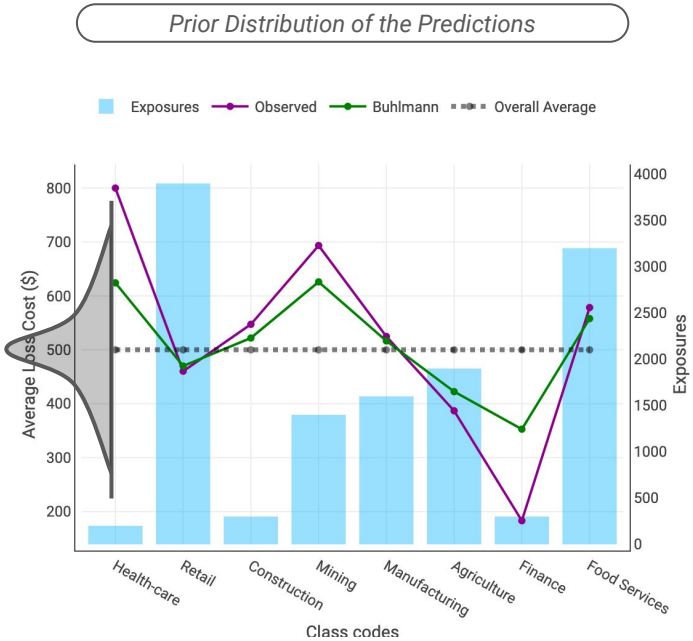
- **The “pure GLM” predictions, centered on the observed values;**
- The “a-priori” distribution of the observations, centered on the grand-average.



What is the idea motivating Credibility?

The Bühlmann credibility creates predictions by mixing two sources of information:

- The “pure GLM” predictions, centered on the observed values;
- **The “a-priori” distribution of the observations, centered on the grand-average.**



What is the idea motivating Credibility?

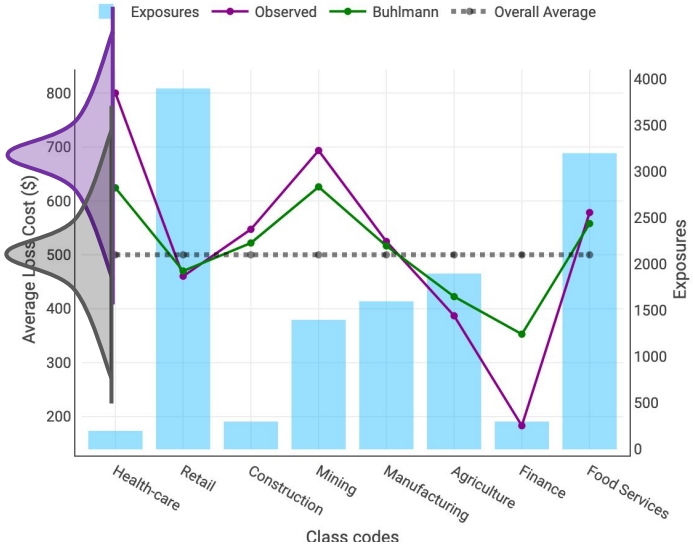
The Bühlmann credibility creates predictions by mixing two sources of informations:

- The “pure GLM” predictions, centered on the observed values;
- The “a-priori” distribution of the observations, centered on the grand-average.

More data means the observed values vary less around the predictions, meaning they can be trusted: a **strong weight** is given to **the observed values**.

Less data means the observed values vary a lot around the predictions, meaning they can’t be trusted: a **strong weight** is given to the **a-priori (grand average)**.

Mixing the two distributions



Quick Reminder… What is Credibility

"Credibility, simply put, is the weighting together of different estimates to come up with a combined estimate."

Foundations of Casualty Actuarial Science

When the volume of data is not enough to accurately estimate the losses, Credibility methodologies provide ways to **complement the observed experience with additional information**.

The Credibility formula is:

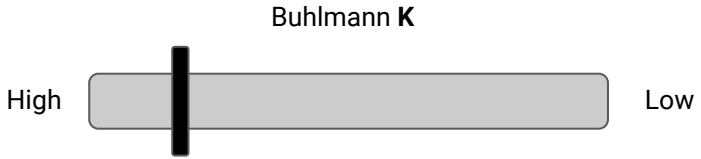
Estimate = Z * **Observed Experience** + (1 – Z) * Complement of Credibility

where the Credibility factor **Z** is a number between 0 and 1.

For example, in Bühlmann Credibility.

$$Z = \frac{n}{n + K}$$

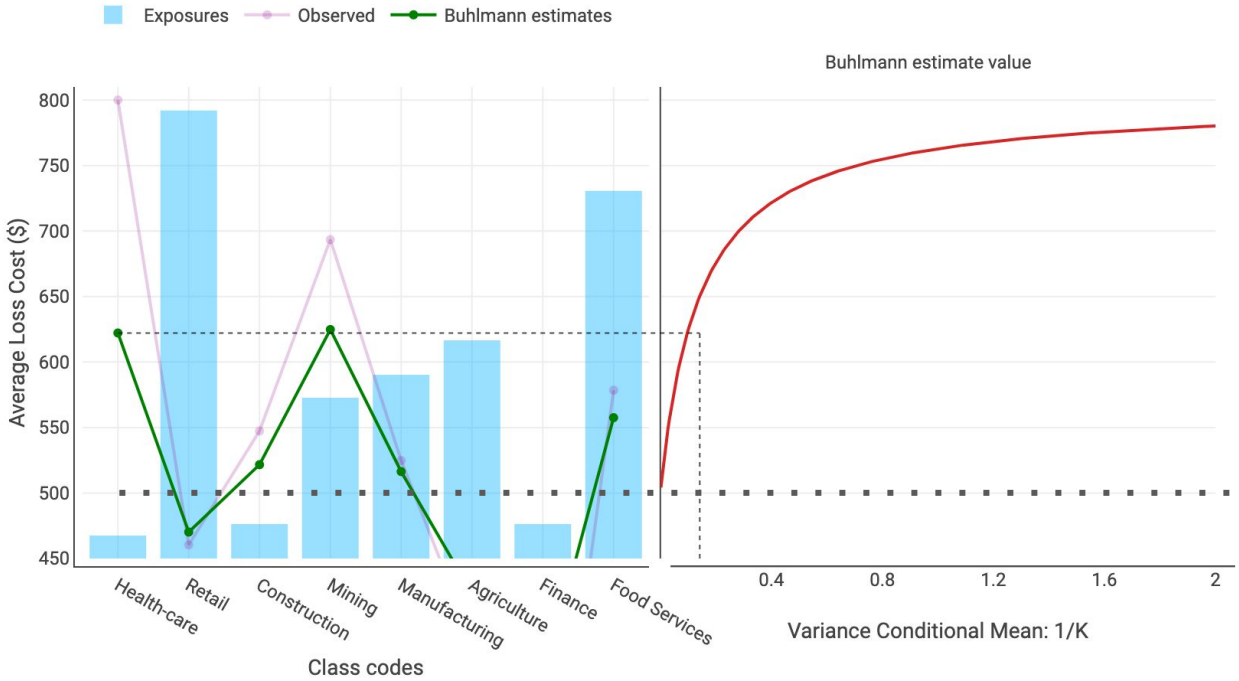
Example: Health Care estimate



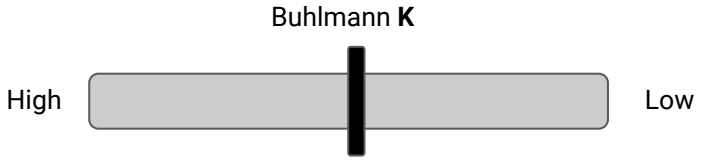
Large K (**low credibility**)

Weak information on the predictions can be derived from the observations (the distributions of the observations around the prediction has a large variance).

Predictions are **close to the overall average**.



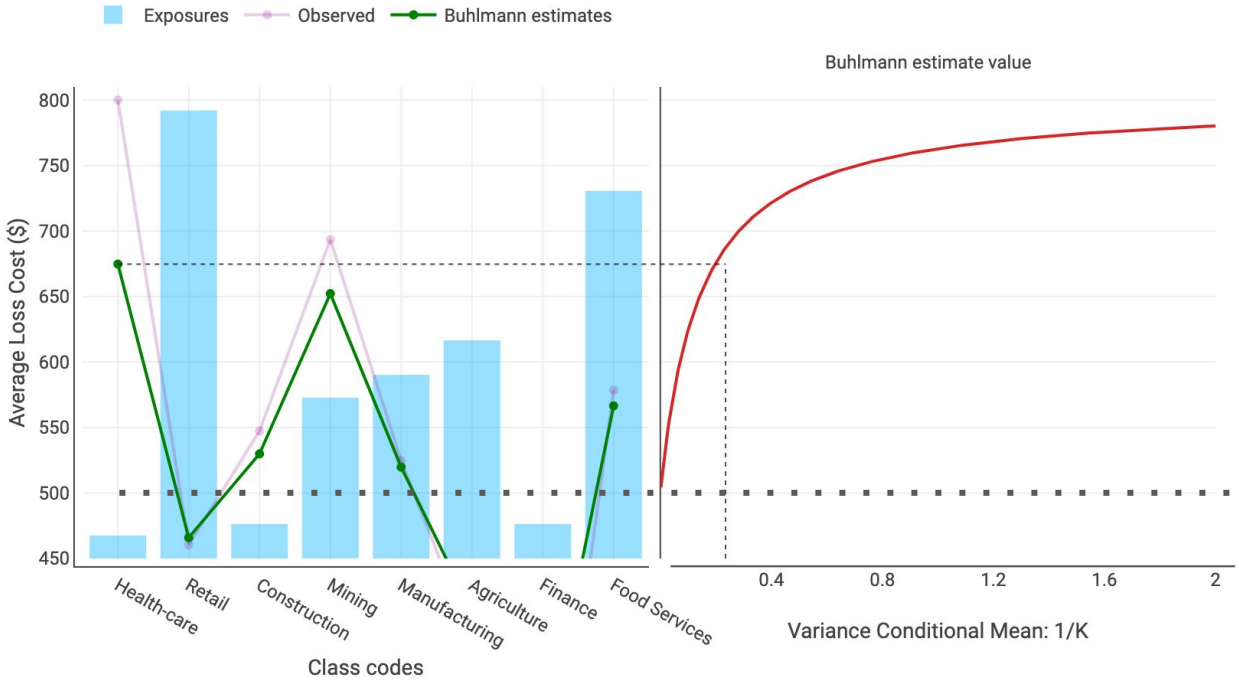
Example: Health Care estimate



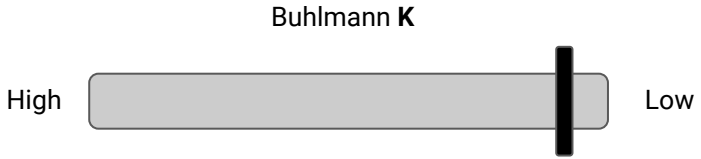
Medium K (**intermediate credibility**)

Intermediate information on the predictions can be derived from the observation (the distributions of the observations around the prediction has a medium variance).

Predictions are **between the overall average and the observations**.



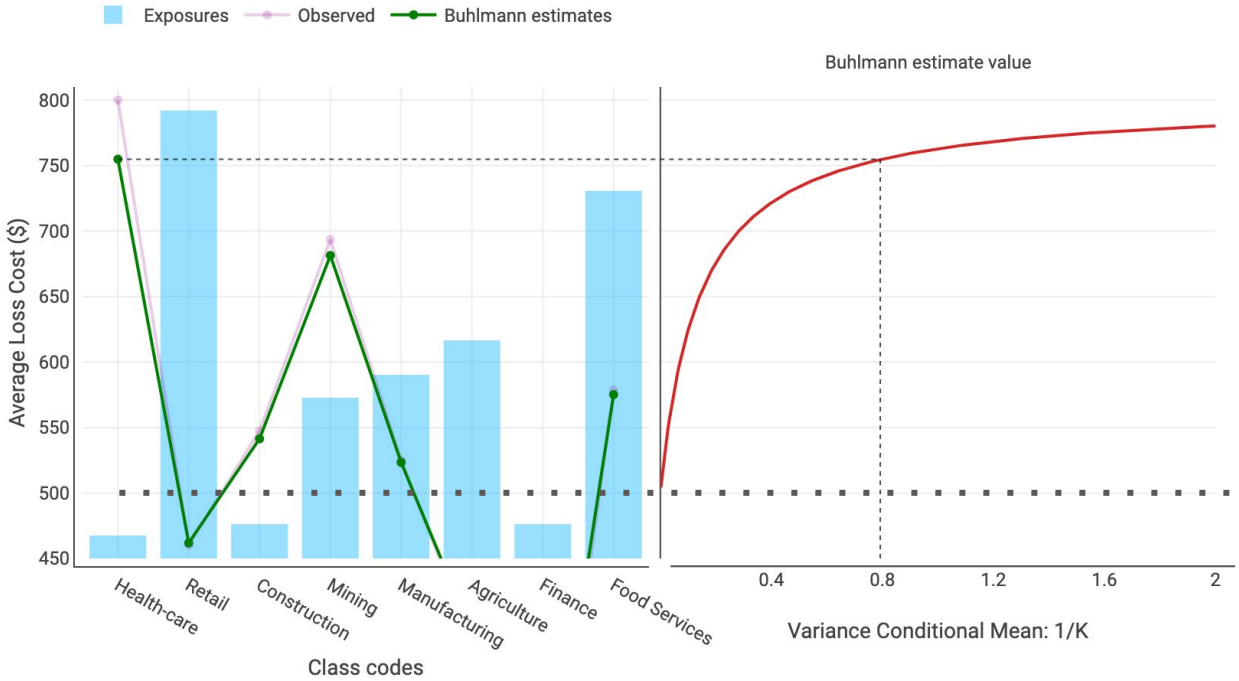
Example: Health Care estimate



Small K (**strong credibility**)

Strong information on the predictions can be derived from the observation (the distributions of the observations around the prediction has a small variance).

Predictions are **close to the observations.**



Strengths & limits of Bühlmann Credibility

This approach has also well-documented strengths & limits:

✅ It allows to leverage all the available data;

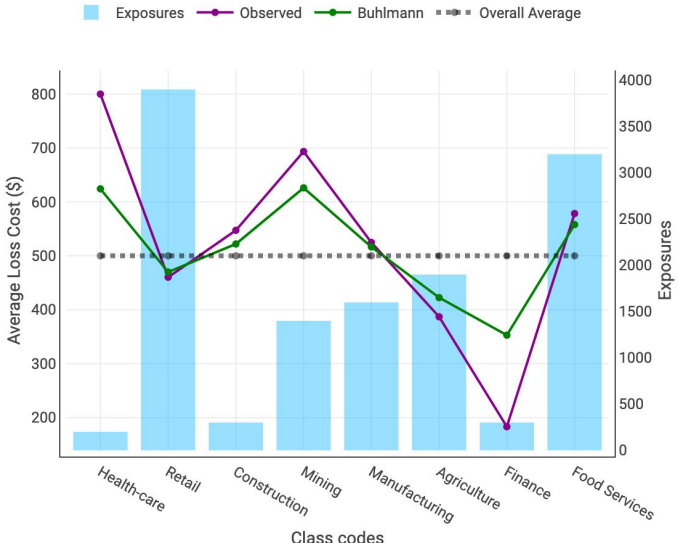
✅ It is very frequently used and widely accepted;

✅ It relies on very classic statistics;

❌ Does not generalize to multiple variables - multivariate modeling

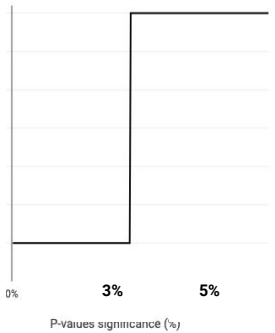
⚠️ It does not select non-significant effects

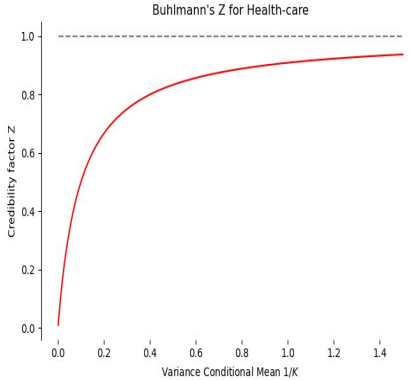
Credibility-based Predictions



Comparison

	P-value significance	Buhlmann Credibility
Set coefficients of low-exposure segments at zero	Selection of effects	No selection of effects
Shrink low-exposure segments	No	This allows to tolerate segments with limited (yet usable) data
Work for multivariate models	Yes	No





Enriching the GLM framework

AKUR8

Multivariate Credibility

The simple GLM estimates are computed as the solution of

$$\beta^* = Argmax\ Likelihood(Obs., \beta)$$

When we apply the same assumptions in Buhlmann Credibility for a multivariate gaussian model, we obtain a Penalized GLM model (Ridge):

$$\beta^* = Argmax\ LogLikelihood(Obs., \beta) - \lambda\, \beta^2$$

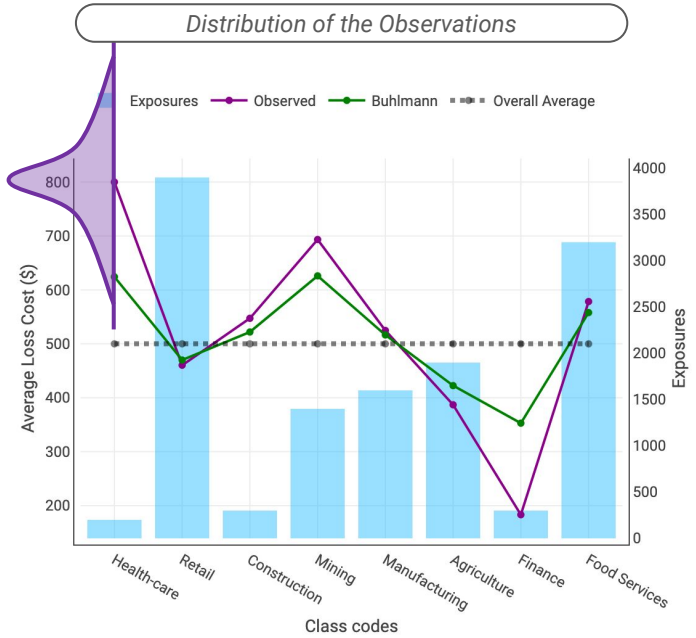
[1] Jewell, W. (1974). Credible Means are exact Bayesian for Exponential Families. *ASTIN Bulletin: The Journal of the IAA, 8*(1), 77-90.

Buhlmann model in multivariate framework

GLM coefficients are the **maximum of likelihood** (probability of observing the data, given the model):

$$\beta^* = Argmax\ Likelihood(Obs., \beta)$$

The probability of observations is displayed in purple on the right.



Buhlmann model in multivariate framework

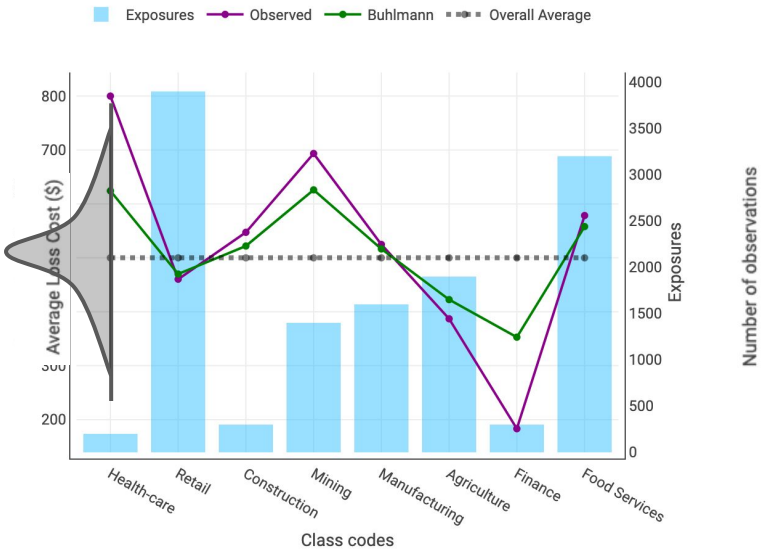
Like for Credibility, Penalized Regressions **integrate another prior hypothesis**.

But this time, **the prior hypothesis is applied directly on the coefficient** values: we integrate a probability for different values of the coefficients.

For instance, in the Ridge-regression framework, we assume coefficients follow a normal distribution:

$$\beta \sim N\left(0, {}^{1}/_{\lambda}\right)$$

Prior Distribution of the Coefficients



The Penalized GLM Formula

This prior is visible in the maximum of likelihood definition:

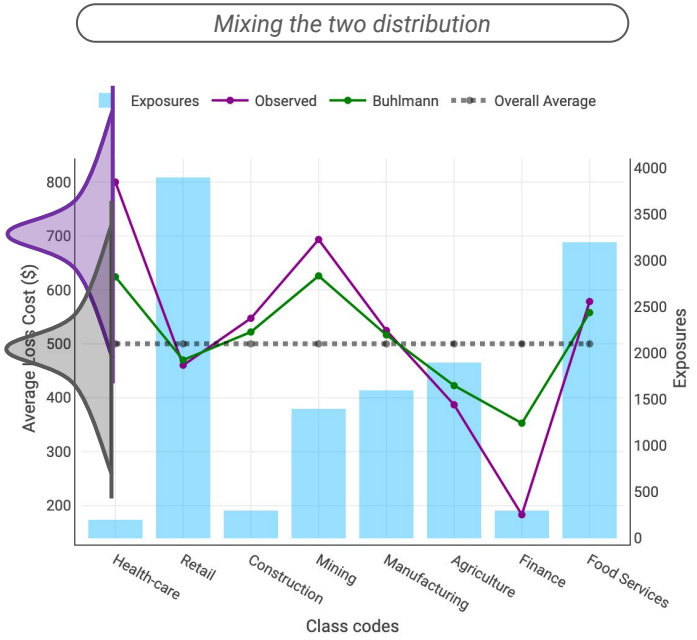
$$\beta^* = Argmax\ Likelihood(Obs., \beta) \times \alpha\ e^{\frac{-\beta^2}{1/\lambda^2}}$$

Which leads to the penalized GLM formula:

$$\beta^* = Argmax\ LogLikelihood(Obs., \beta) - \lambda\ \beta^2$$

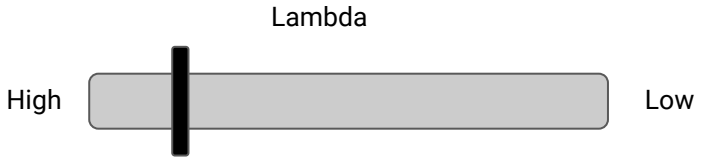
The formula above is a Penalized GLM known as Ridge.

- **Ridge regression is exactly equal to Buhlmann estimates under Gaussian assumption;**
- **In that case, lambda is exactly equal to the K parameter.**



Example: Health Care estimate

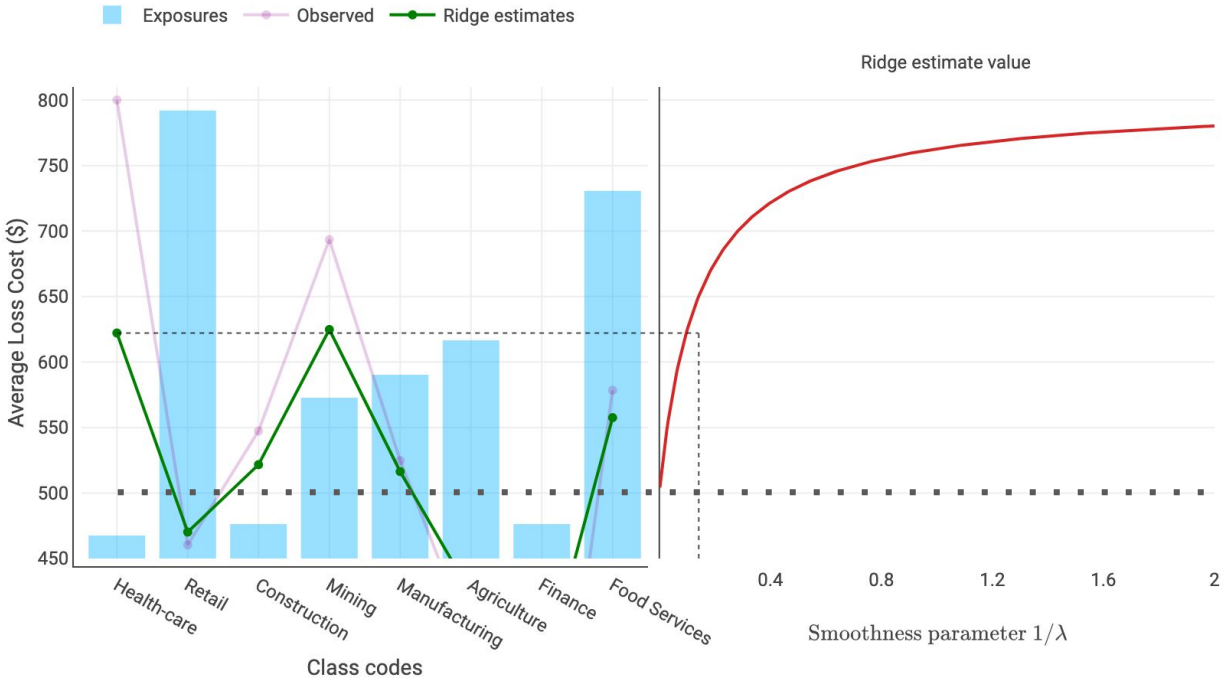
Ridge Regression is a Credibility Procedure



Large λ (large penalty)

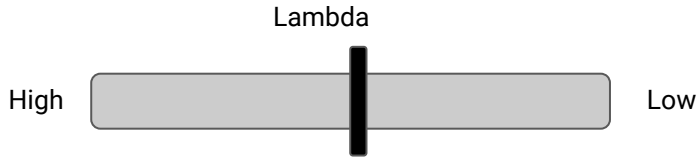
Strong prior on the coefficient (the prior distribution has a small variance).

Coefficients and predictions are **close to the overall average**.



Example: Health Care estimate

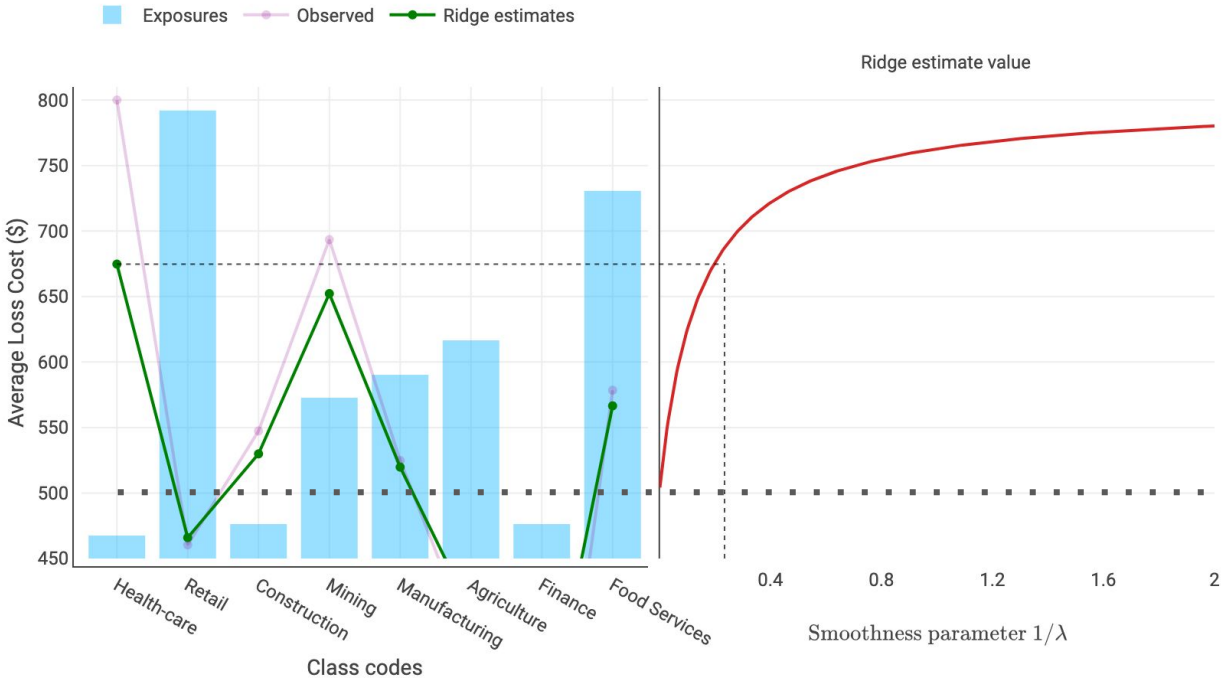
Ridge Regression is a Credibility Procedure



Medium λ (medium penalty)

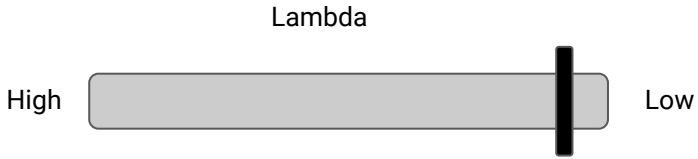
Intermediate prior on the coefficient (the prior distribution has a small variance).

Coefficients and predictions are **further to the overall average**.



Example: Health Care estimate

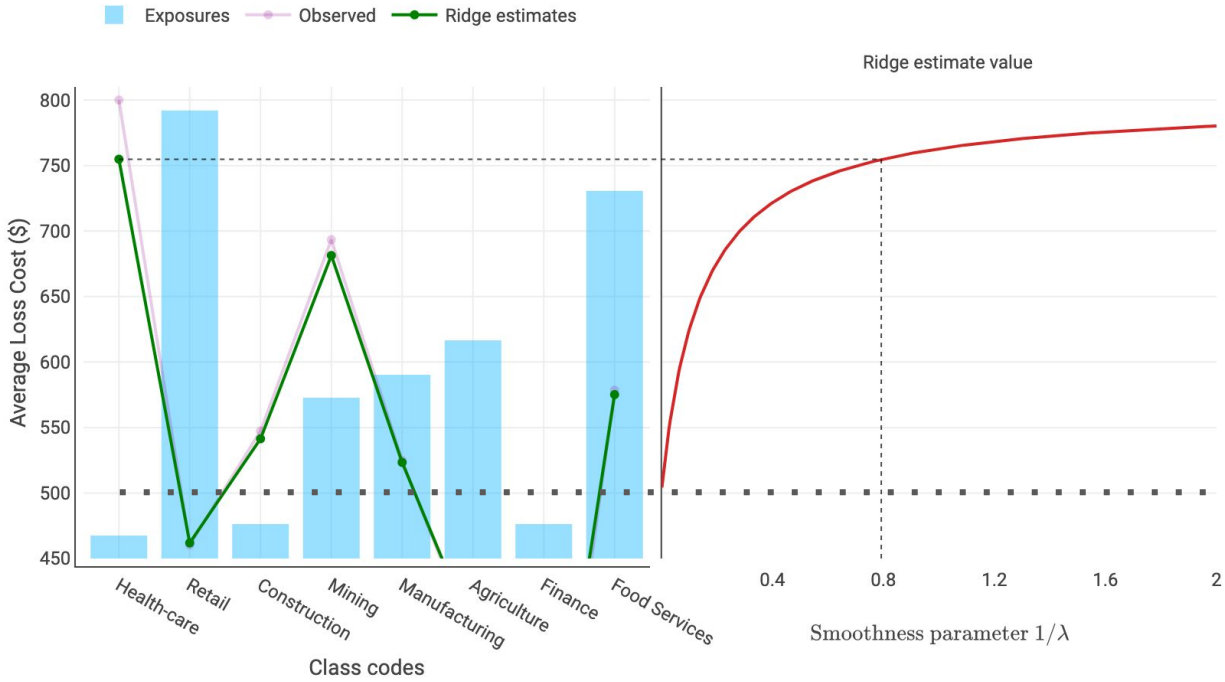
Ridge Regression is a Credibility Procedure



Small λ (small penalty)

Weak prior on the coefficient (the prior distribution has a large variance).

Coefficients and predictions are **close to the observed value**.

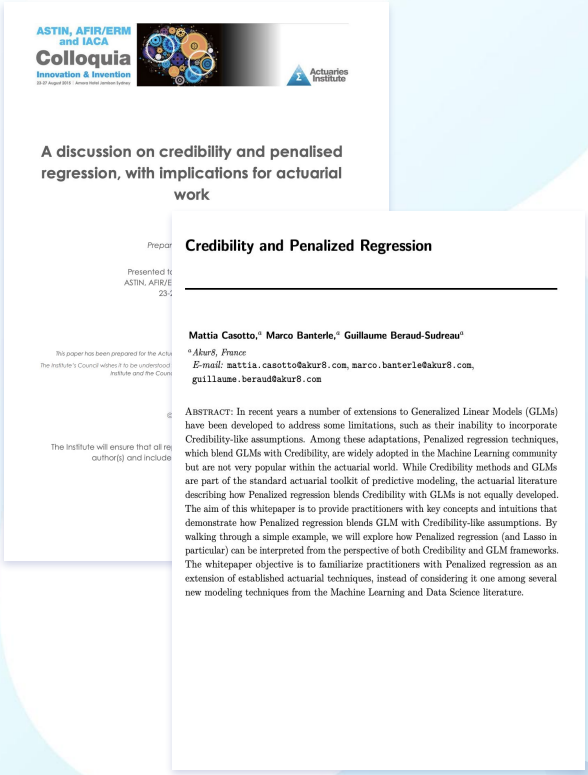


Blending GLM with Credibility

Penalized GLMs share the same properties as **Credibility** in the following ways:

1. Both **shrink** GLM estimates toward the complement of Credibility (grand average);
2. Both apply **more shrinkage** to segments with **low volume** of data / credibility
3. The Credibility approach can be **applied to predictions** (or one variable). The ridge regression can be applied to **all variables simultaneously**.

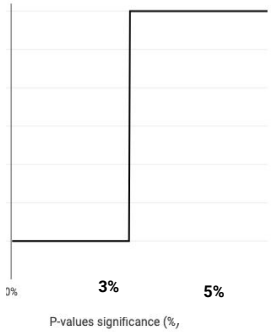
Penalized Regression can meet the definition of a credibility procedure according to the definition provided in ASOP 25: Credibility Procedures.

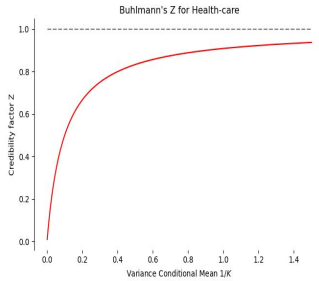


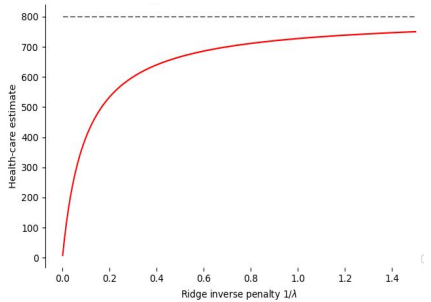
Comparing different techniques

Penalized GLM enhances GLMs by including credibility considerations

	P-value significance	Credibility	Ridge Regression
Set coefficients of low-exposure segments at zero	Selection of effects	No selection of effects	
Shrink low-exposure segments	No	This allows to tolerate segments with limited (yet usable) data	
Work for multivariate models	Yes	No	Yes







Credibility and segment selection

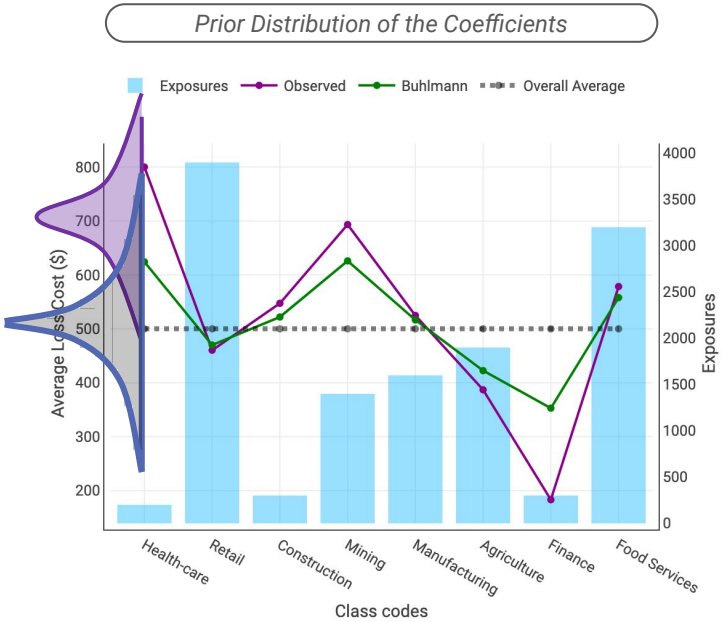
Lasso Penalization can be used as a Credibility Procedure as well

The Lasso GLM is a Penalized GLM that is able to both shrink the estimates (as in Credibility) and to set to zero those coefficients which are not material to the model (as in level selection and p-values).

The formula for the Lasso is

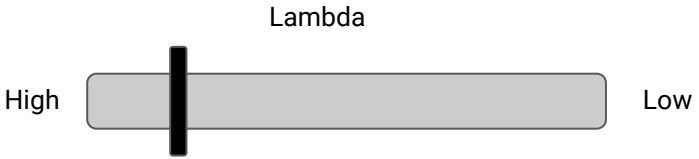
$$\beta^* = Argmax\ LogLikelihood(Obs., \beta) - \lambda|\beta|$$

The formula corresponds to a Laplace prior, which is a very “pointy”, meaning that coefficients have a high probability of being exactly zero.



Impact of smoothness to Lasso estimates

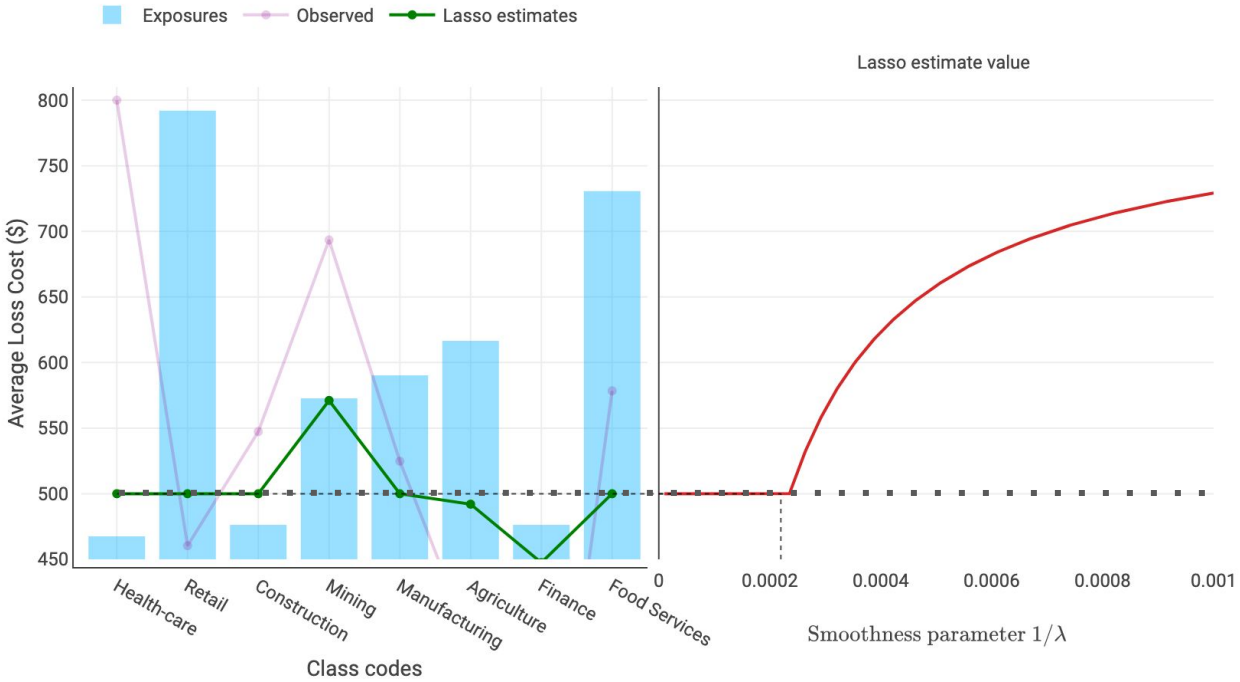
Workers Compensation example



Large λ (large penalty)

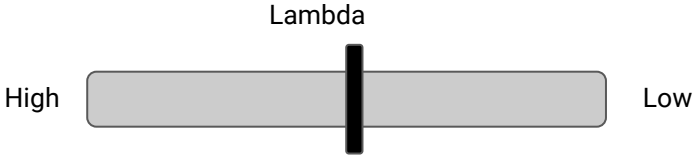
Strong prior on the coefficient (the prior distribution has a small variance).

Coefficients and predictions are **close to the overall average**.



Impact of smoothness to Lasso estimates

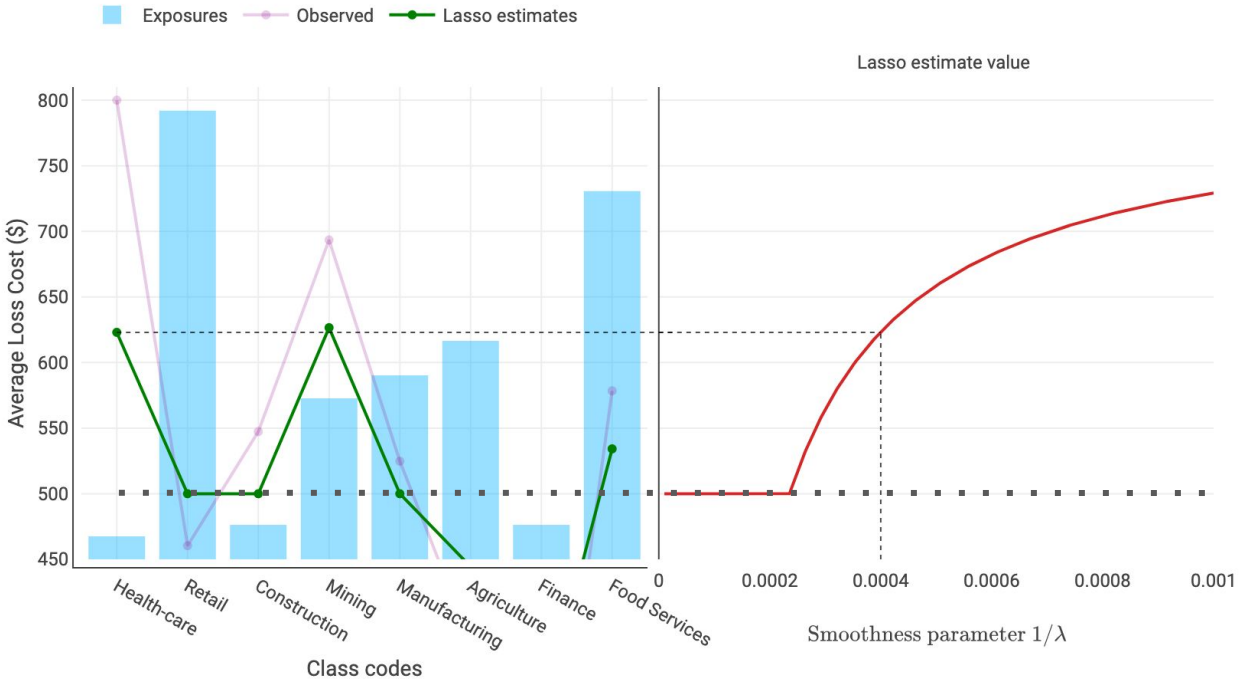
Workers Compensation example



Medium λ (medium penalty)

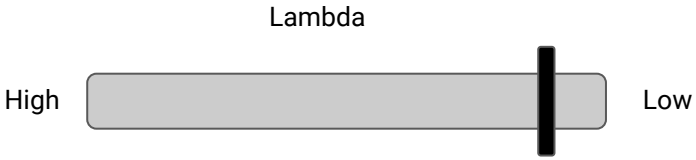
Intermediate prior on the coefficient (the prior distribution has a small variance)

Coefficients and predictions are **further to the overall average**.



Impact of smoothness to Lasso estimates

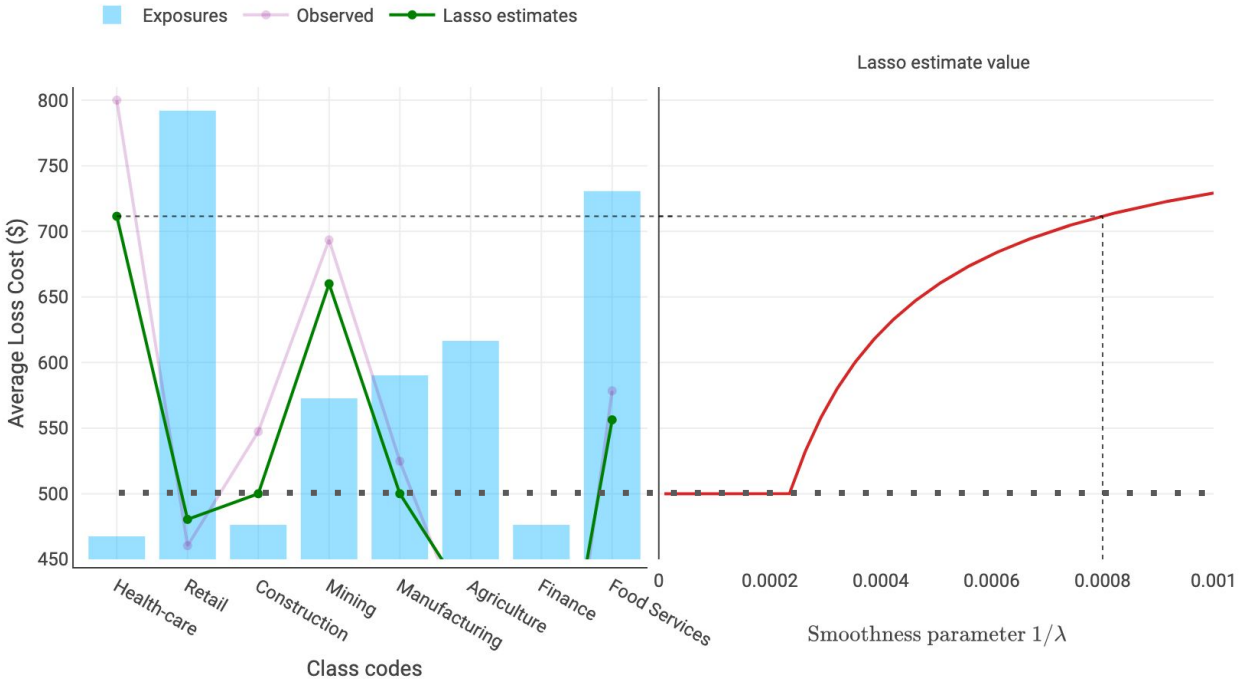
Workers Compensation example



Small λ (small penalty)

Weak prior on the coefficient (the prior distribution has a large variance)

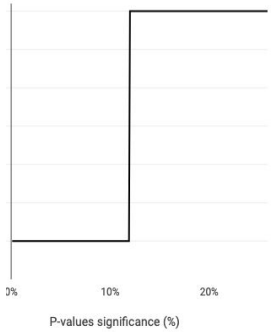
Coefficients and predictions are **close to the observed value**.

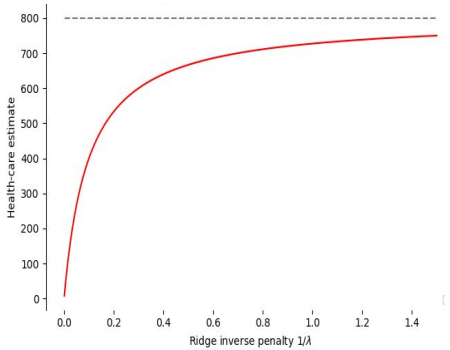


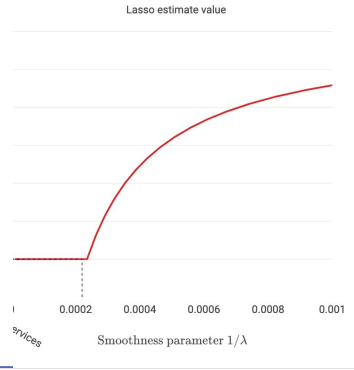
Comparing different techniques

Lasso combines credibility benefits of Ridge and can remove insignificant coefficient

	P-value significance	Ridge GLM	Lasso Regression
Set coefficients of low-exposure segments at zero	Selection of effects	No selection of effects	Selection of effects
Shrink low-exposure segments	No	This allows to tolerate segments with limited (yet usable) data	
Work for multivariate models	Yes		





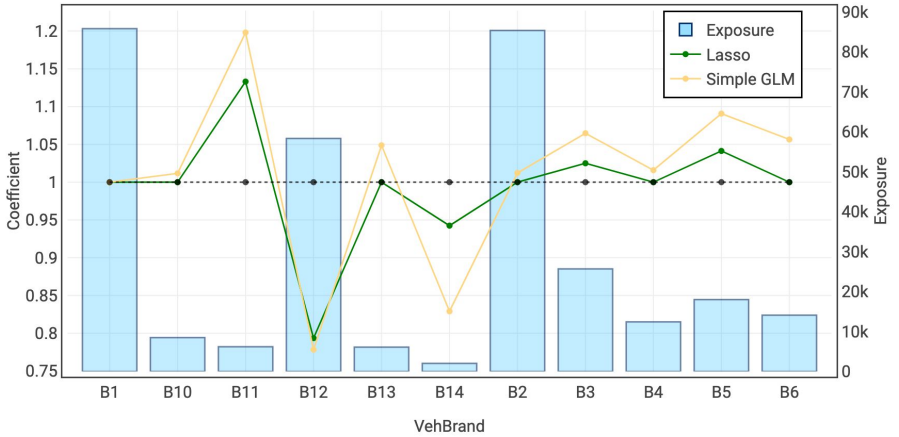


Blending credibility and levels selection

Lasso penalization removes the need for P-value review

In this example, **Lasso Penalization has automatically set levels with GLM P-values above 0.05 to a neutral factor of 1.0.**

Level B14 has a P-value of 0.046 in the GLM. **Lasso Penalization has shrunk this coefficient** to reflect the instability in this level.



VehBrand	GLM Coefficients	Lasso Coefficients	P-Values
B1	1.00000	1.00000	nan
B10	1.01167	1.00000	0.78277
B11	1.19804	1.13311	0.00005
B12	0.77841	0.79358	0.00000
B13	1.04876	1.00000	0.31601
B14	0.82906	0.94246	0.04617
B2	1.01219	1.00000	0.50507
B3	1.06474	1.02499	0.01297
B4	1.01589	1.00000	0.65058
B5	1.09073	1.04132	0.00289
B6	1.05653	1.00000	0.09113

Lasso Regression can simplify model review by applying credibility considerations as well as automating coefficient removal.

Penalized GLMs and GBMs

AKUR8

Ordinal variables and GLMs

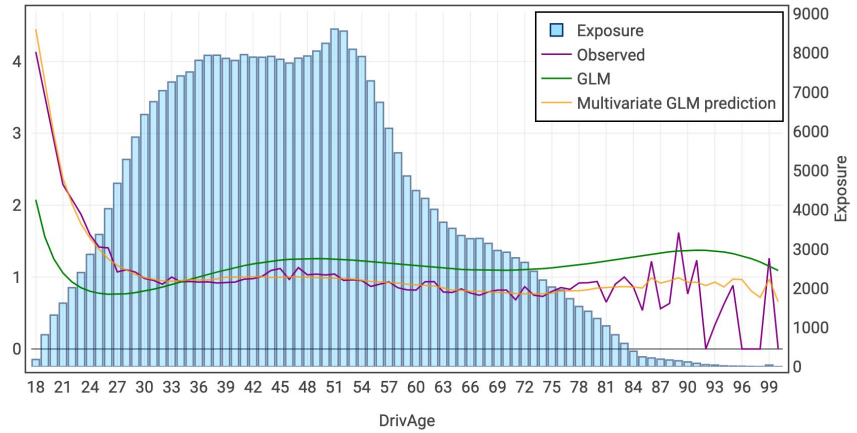
```
ClaimNb ~ DrivAge + I(DrivAge**2) + I(DrivAge**3) + I(DrivAge**4) + log(DrivAge)
```

GLMs can model non-linear relationships via “feature engineering” such as polynomial transformations.

The figure showcases an example of a driverAge GLM modeled using a 4th degree polynomial + a log term. All parameters are statistically significant.

We observe a risk increase for drivers above 80, which may not be credible by the lack of exposures for that segment.

The price increase seem to be an artifact of the polynomial transformation more than material to the data.



Imperfect variable transformations can be significant while performing poorly where exposures are thin.

Polynomial instability

Polynomial transformations may not be appropriate because of the instabilities they exhibit on the tails.

We compare two GLMs whose effects are statistically sound (< 0.5% p-value)

- “With log” - models age with 4th degree polynomial + log
- “Without log” - models age with 4th degree polynomial

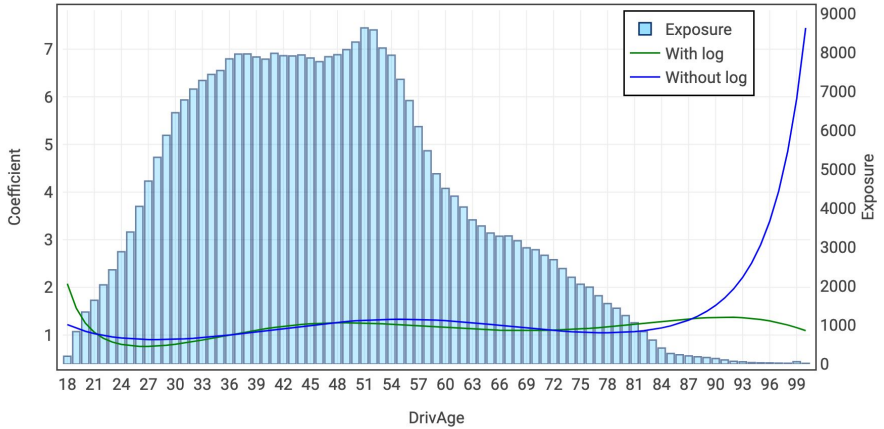
The estimates of the model are wildly different for the older tail - even if both models are statistically sound.

With log - Significant

	P-values
Intercept	0.00000
np.log(DrivAge)	0.00000
DrivAge	0.00000
I(DrivAge ** 2)	0.00000
I(DrivAge ** 3)	0.00000
I(DrivAge ** 4)	0.00000

Without - Still Significant!

	P-values
Intercept	0.00000
DrivAge	0.00000
I(DrivAge ** 2)	0.00000
I(DrivAge ** 3)	0.00000
I(DrivAge ** 4)	0.00000

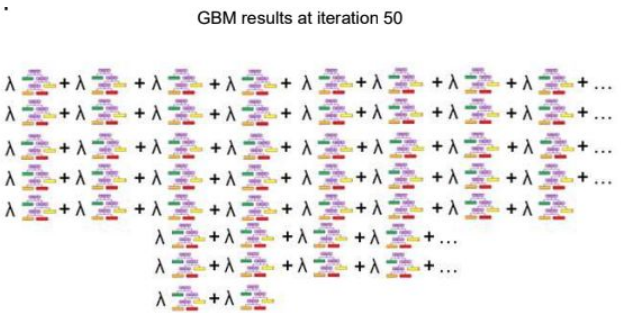


Gradient Boosting Machines

Gradient Boosting Machines (GBMs) are a tree-based modeling technique popular in the Machine Learning community

Benefits

1. GBMs natively fit non-linearities and require no previous data transformations
2. GBMs can fit any arbitrary amount of interactions



Drawbacks

1. GBMs are **not transparent** and **hard to audit**
2. It is not possible to audit the parameters but approximations of impacts of model in the data (Partial Dependency Plot, ICE plots, Lime, Shap Values,....)
3. They are a tree-based methodology: their mathematical framework is different than standard GLMs

In the following slides, we will inspect how GBMs are able to detect non-linear relationships.

By developing these intuitions, we can build a transparent penalized GLM that can fit non-linear relationships natively and robustly.

GBM and Ordinal variables

GBMs natively handles **non-linear effects** by combining

1. **Trees**

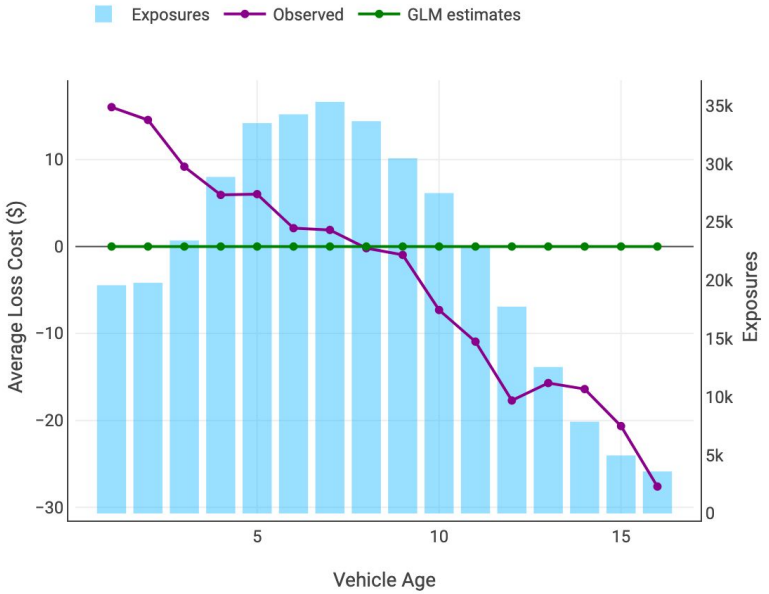
   Detects the location on where to split the ordinal variables in two region

2. Learning Rate

   Allows to incrementally adapt the trees to the signal, making the model ‘smoother’ and more robust to correlations

3. Boosting

   Adaptively learns structure from the residuals / errors



GBM and Ordinal variables

GBMs natively handles **non-linear effects** by combining

1. **Trees**

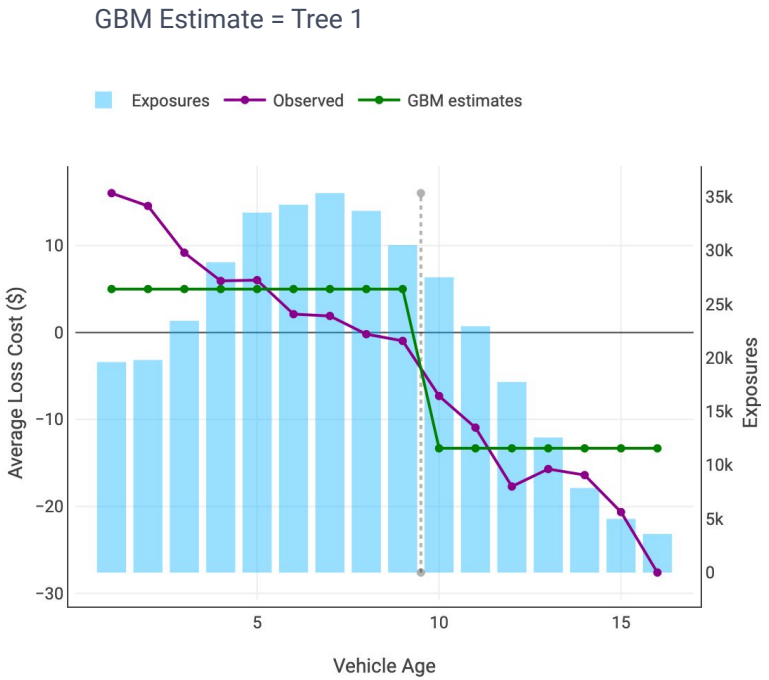
   Detects the location on where to split the ordinal variables in two region

2. Learning Rate

   Allows to incrementally adapt the trees to the signal, making the model ‘smoother’ and more robust to correlations

3. Boosting

   Adaptively learns structure from the residuals / errors



The impact of the Learning Rate

GBMs natively handles **non-linear effects** by combining

1. Trees

   Detects the location on where to split the ordinal variables in two region

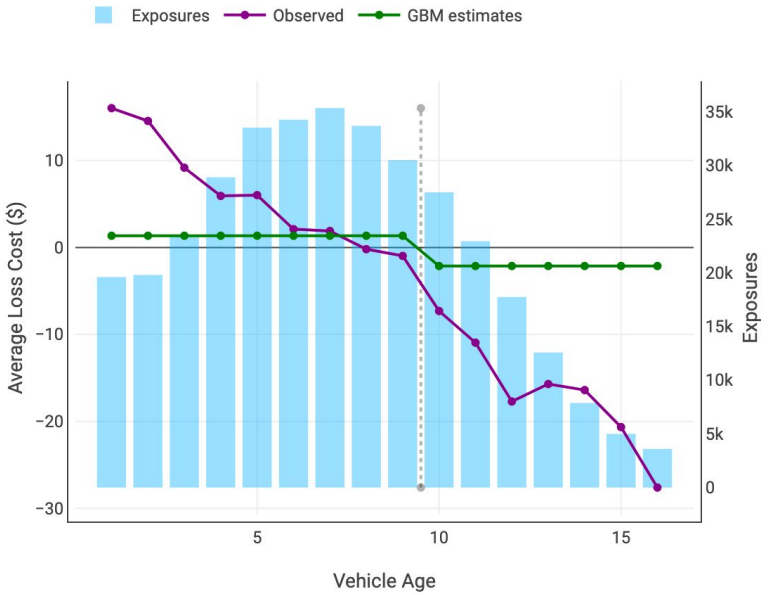
2. **Learning Rate**

   Allows to incrementally adapt the trees to the signal, making the model 'smoother' and more robust to correlations

3. Boosting

   Adaptively learns structure from the residuals / errors

GBM Estimate = **0.1** * Tree 1



1.

The impact of the Learning Rate

GBMs natively handles **non-linear effects** by combining

1. Trees

   Detects the location on where to split the ordinal variables in two region

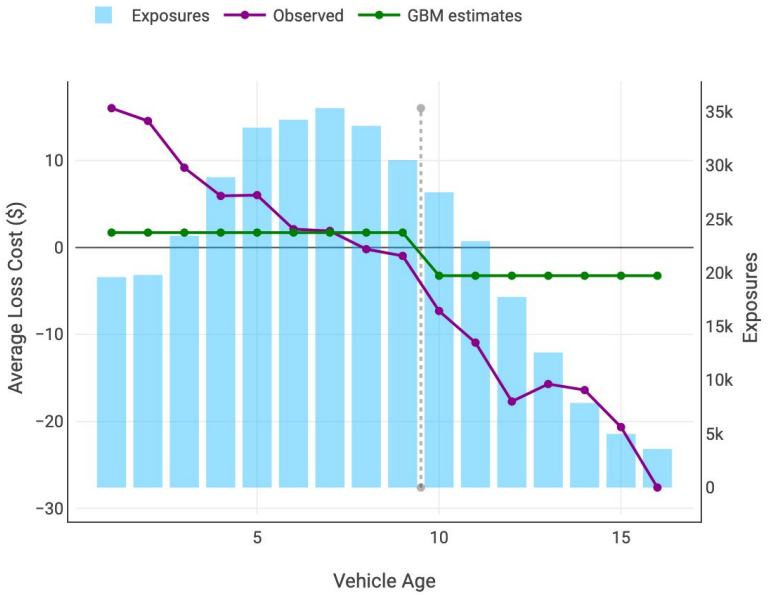
2. Learning Rate

   Allows to incrementally adapt the trees to the signal, making the model ‘smoother’ and more robust to correlations

3. **Boosting**

   Adaptively learns structure from the residuals / errors

GBM Estimate = **0.5** * Tree 1 + **0.5** * Tree 2



The impact of the Learning Rate

GBMs natively handles **non-linear effects** by combining

1. Trees

   Detects the location on where to split the ordinal variables in two region

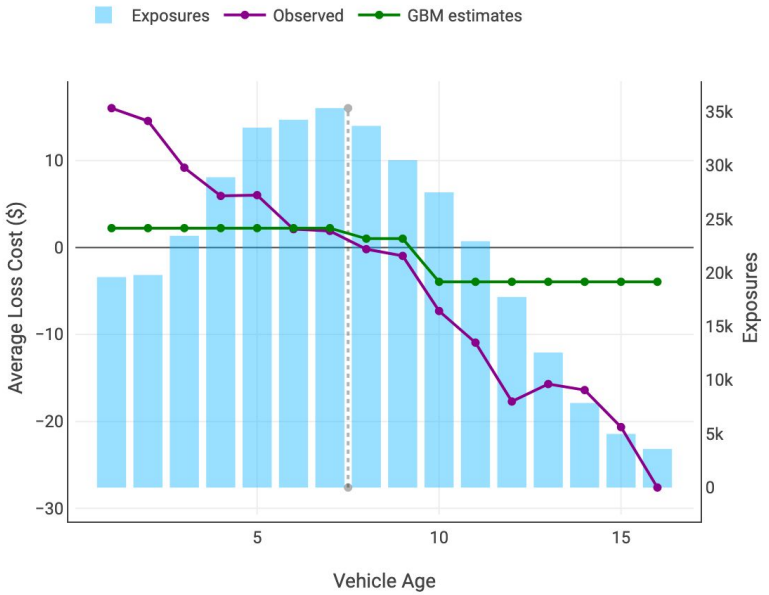
2. Learning Rate

   Allows to incrementally adapt the trees to the signal, making the model ‘smoother’ and more robust to correlations

3. **Boosting**

   Adaptively learns structure from the residuals / errors

GBM Estimate = **0.5** * Tree 1 +... + **0.5** * Tree 3



The impact of the Learning Rate

GBMs natively handles **non-linear effects** by combining

1. Trees

   Detects the location on where to split the ordinal variables in two region

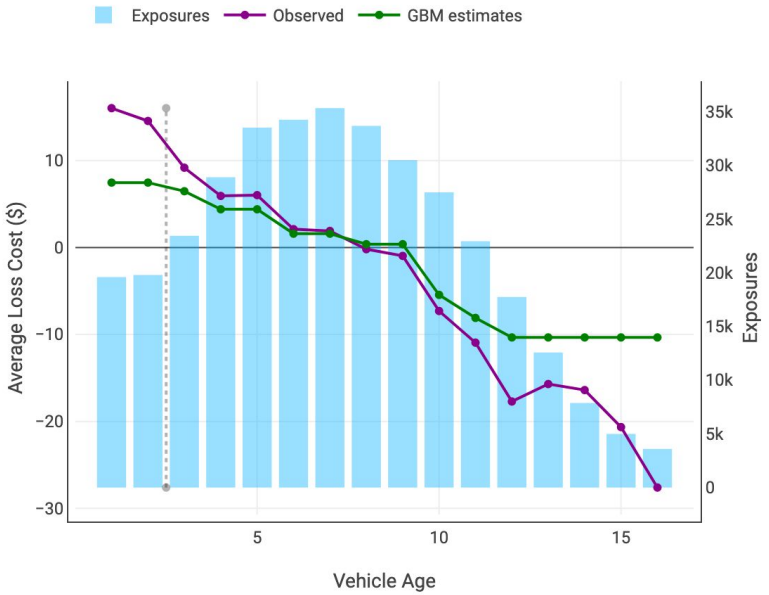
2. Learning Rate

   Allows to incrementally adapt the trees to the signal, making the model ‘smoother’ and more robust to correlations

3. **Boosting**

   Adaptively learns structure from the residuals / errors

GBM Estimate = **0.5** * Tree 1 +... + **0.5** * Tree 11



How GBMs ‘learn’ ordinal variables

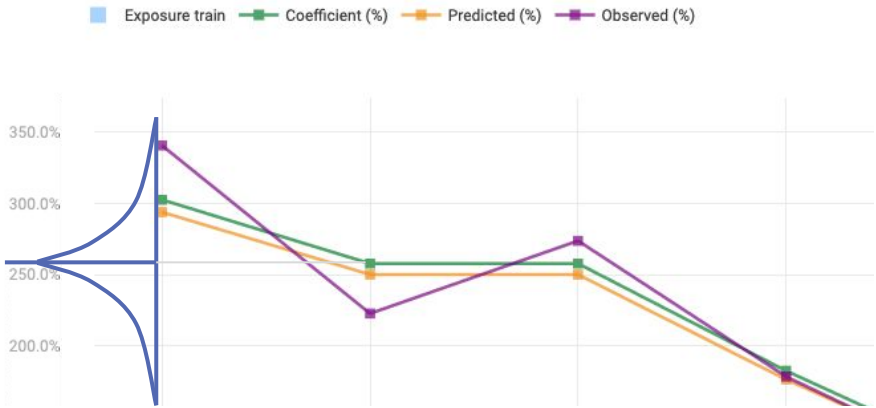
GBM learn non-linearities by **adaptively grouping the variable,** based on the underlying signal**.**

Penalized regression can replicate this structure by using an appropriate **prior distribution** (or **penalty**): the **derivative Lasso**.

Creating new Priors and Penalties

Grouping is statistically equivalent to the assumption that the coefficients of two consecutive levels:

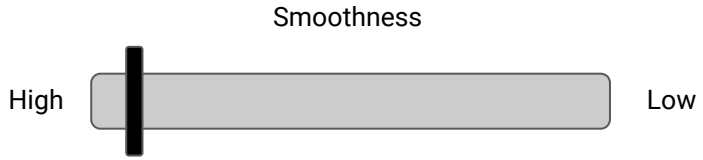
- **Are more likely to be close than far apart** if their difference is statistically significant.
- Or **have the same coefficients** if the levels do not have enough data to have a statistically significant difference.This behavior can be modeled by assuming that the **derivative of the (ordinal) variable follows a Laplace distribution**, leading to the **Derivative Lasso** formulation:



$$\beta^* = Argmax_\beta \; LL(x, y, \beta) - \lambda \, |\beta_i - \beta_{i+1}|$$

This formulation continues to maximize the likelihood like a traditional GLM, but with additional credibility considerations.

Lasso and Ordinal variables

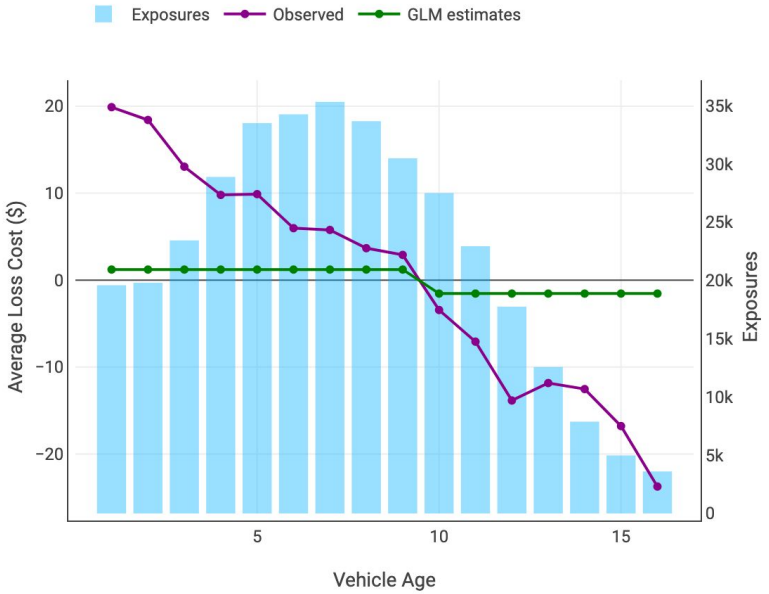


Under these **"Lasso"** assumption on the **derivative**, penalized regression can **natively incorporate non-linear effects**.

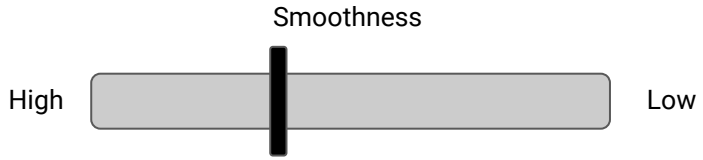
Furthermore, the convergence result between GBMs and Lasso is still valid.

To control the training error and ability to generalise:

- Penalized Regression require the definition of a **single credibility-based parameter**: the **smoothness**
- GBMs require to determine the combination of **several parameters:**
  - **number of trees**
  - **learning rate**
  - and other tree-related parameters



Lasso and Ordinal variables

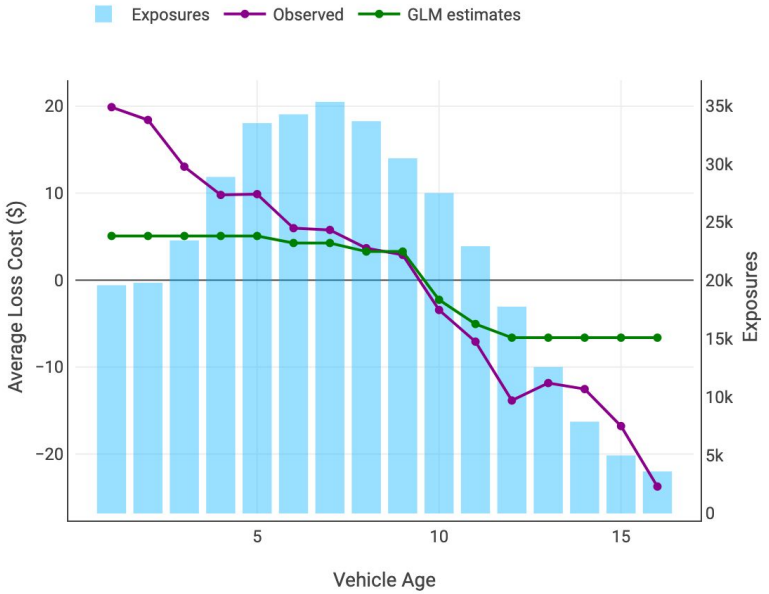


Under these **"Lasso"** assumption on the **derivative**, penalized regression can **natively incorporate non-linear effects**.

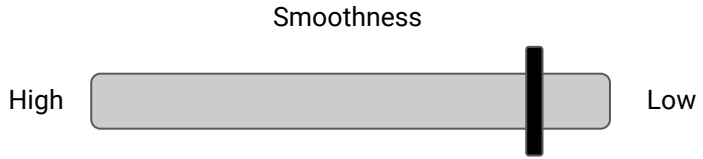
Furthermore, the convergence result between GBMs and Lasso is still valid.

To control the training error and ability to generalise:

- Penalized Regression require the definition of a **single credibility-based parameter**: the **smoothness**
- GBMs require to determine the combination of **several parameters:**
  - **number of trees**
  - **learning rate**
  - and other tree-related parameters



Lasso and Ordinal variables

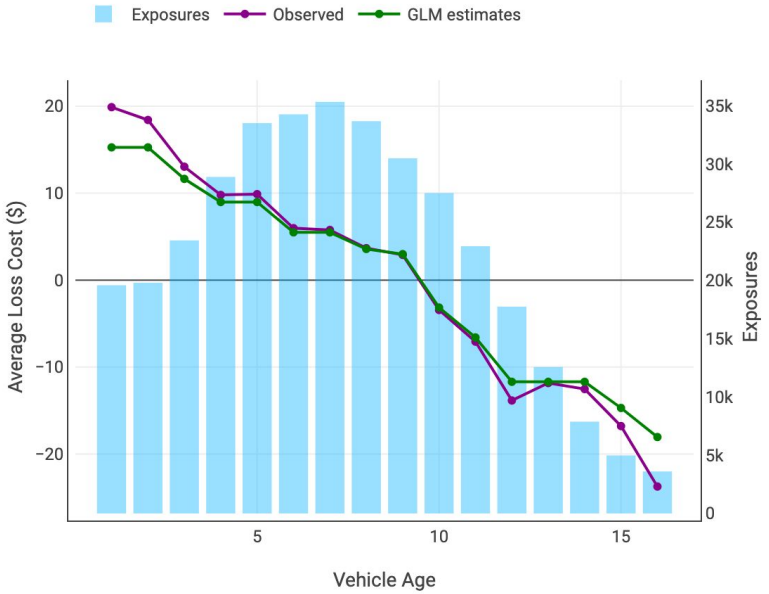


Under these **"Lasso"** assumption on the **derivative**, penalized regression can **natively incorporate non-linear effects**.

Furthermore, the convergence result between GBMs and Lasso is still valid.

To control the training error and ability to generalise:

- Penalized Regression require the definition of a **single credibility-based parameter**: the **smoothness**
- GBMs require to determine the combination of **several parameters:**
  - **number of trees**
  - **learning rate**
  - and other tree-related parameters



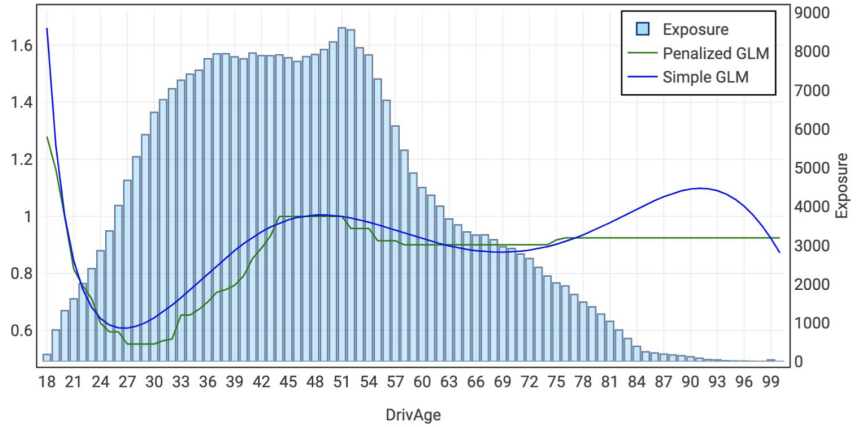
Derivative Lasso and GLM

Via **adaptive grouping**, the derivative Lasso is able to model only **significant non-linearities**.

By its nature, it is not showing instabilities at the older tail of the distribution, as in polynomial modeling.

In this specific example, the resulting rate should be fairer as the older segment is charged constantly accordingly to the observed experience, and not as an artefact of the polynomial modeling.

Furthermore, as the derivative Lasso is a penalized GLM, it inherits the ability to blend a model with credibility.



Conclusion

Penalized GLM offers a **flexible and theoretically sound** framework to tackle and address the GLM's drawbacks.

It does so in an **accessible** way:

- Penalized GLM require the choice of **only one parameter: the smoothness (lambda)**
    - Relates to known credibility techniques
    - Derivative Lasso adaptive grouping can represent non-linearities, without choosing manual transformations whose decisions may be biased by the modeler.
- Penalized GLMs share the **same assumptions and the same output as a GLM**
    - The same visual analysis to evaluate the quality of a GLM helds in a Penalized GLM

Conclusion pt.2

1. Penalized Regression enhances GLM methodology by **natively including credibility considerations**.

2. Lasso penalization simplifies model review by **automatically removing insignificant variables** and shrinking variables that are significant but cannot be fully trusted.

3. The Derivative Lasso framework further enhances Lasso Penalization by borrowing techniques from GBMs to **automatically fit non-linearities** in a data-driven and transparent methodology.

THANKS

9 rue Fortuny, 75017 Paris
FRANCE

AKUR8