S. I. A. s.r.l.
SVILUPPO INIZIATIVE ATTUARIALI

Regressione Penalizzata e Credibilità nei GLM

2025-06-26

Leonardo Stincone

Actuarial Data Scientist

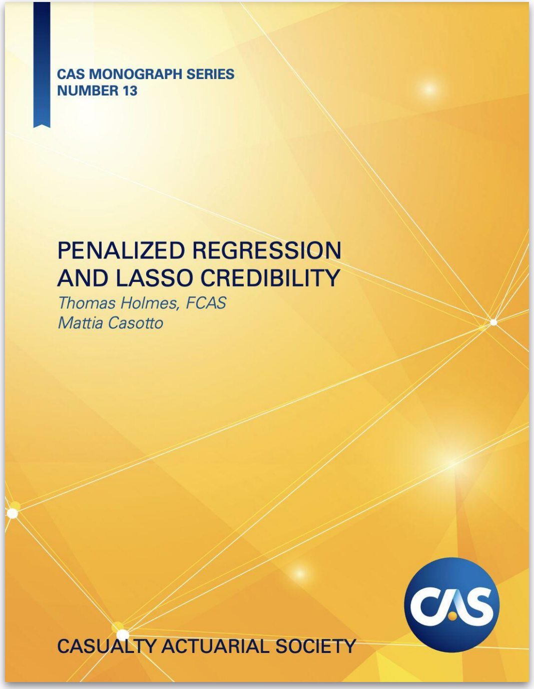
Leonardo.Stincone@akur8.com
linkedin.com/in/leonardo-stincone/

Mattia Casotto

Head of Product USA

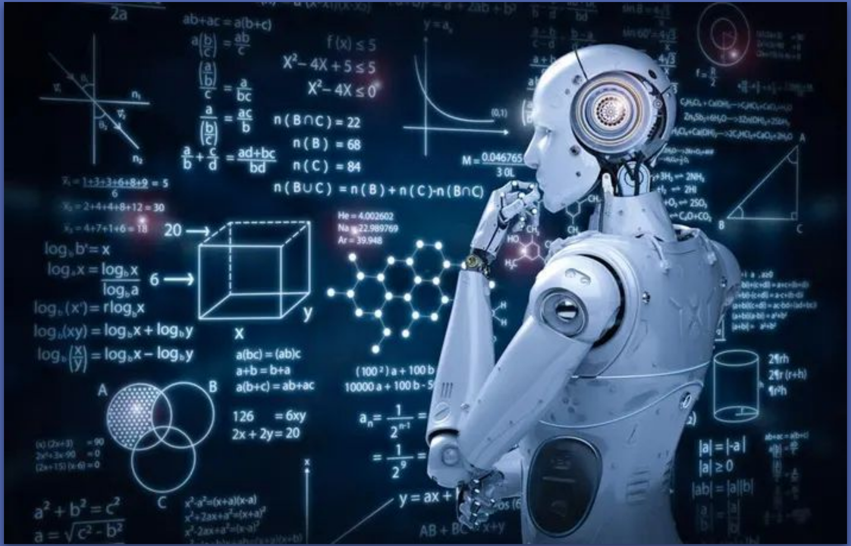
Mattia.Casotto@akur8.com
linkedin.com/in/mattia-casotto/

Have a look at our latest CAS Monograph



Thomas Holmes
Chief Actuary USA

Mattia Casotto
Head of Product USA



Penalized Regression and Credibility in GLMs

Agenda

1. P&C Pricing: Introduction and different approaches
2. Generalized Linear Models in Pricing: Main challenges
3. Penalized Regression and Credibility: Definition, interpretation and key benefits
4. Conclusions and Q&A

Let’s get to know each others

Kahoot!

P&C Pricing: Introduction

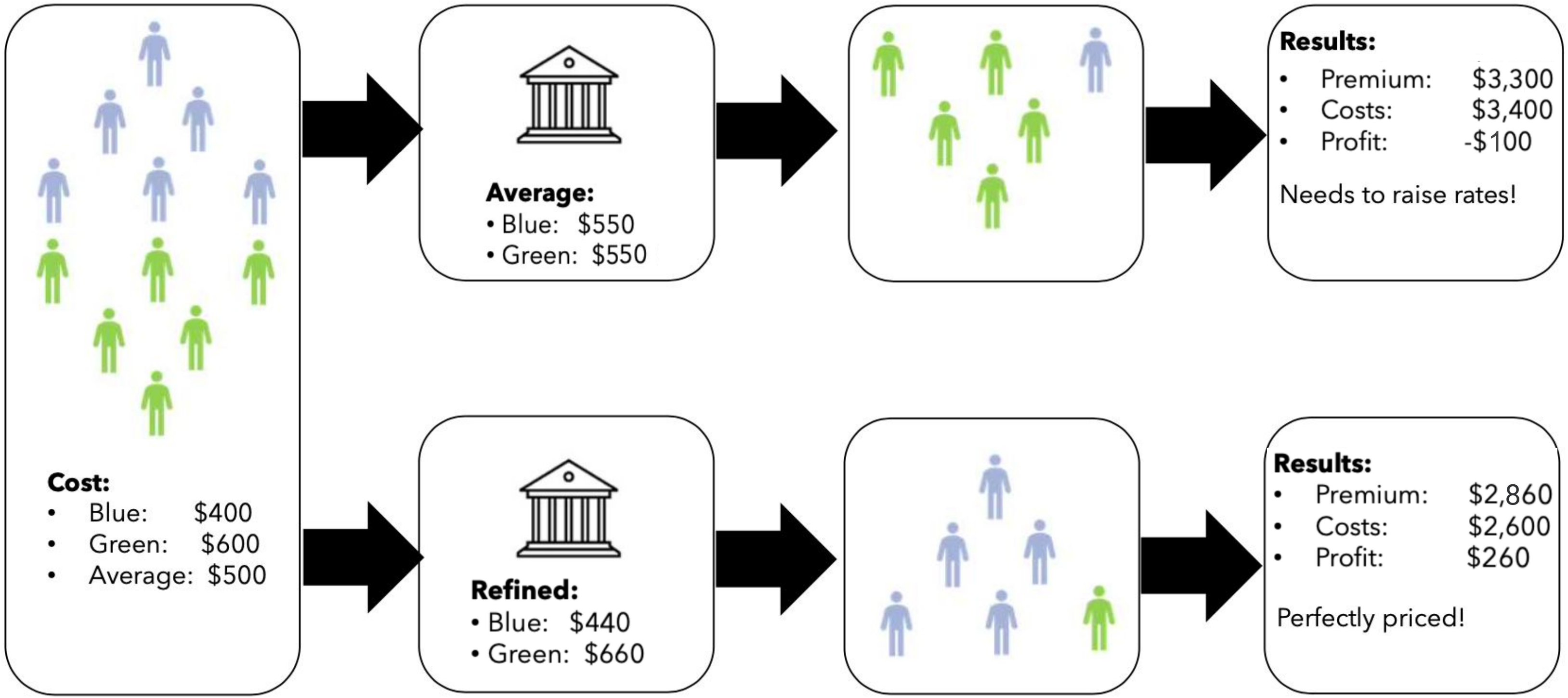
P&C Pricing

Price Personalization: In Insurance typically different policyholders pay different prices

		Age	Location	Car	Price
	Alberto	25	Rome		400 €
	Barbara	50	Milan		200 €
	Carlo	40	Naples		500 €

P&C Pricing

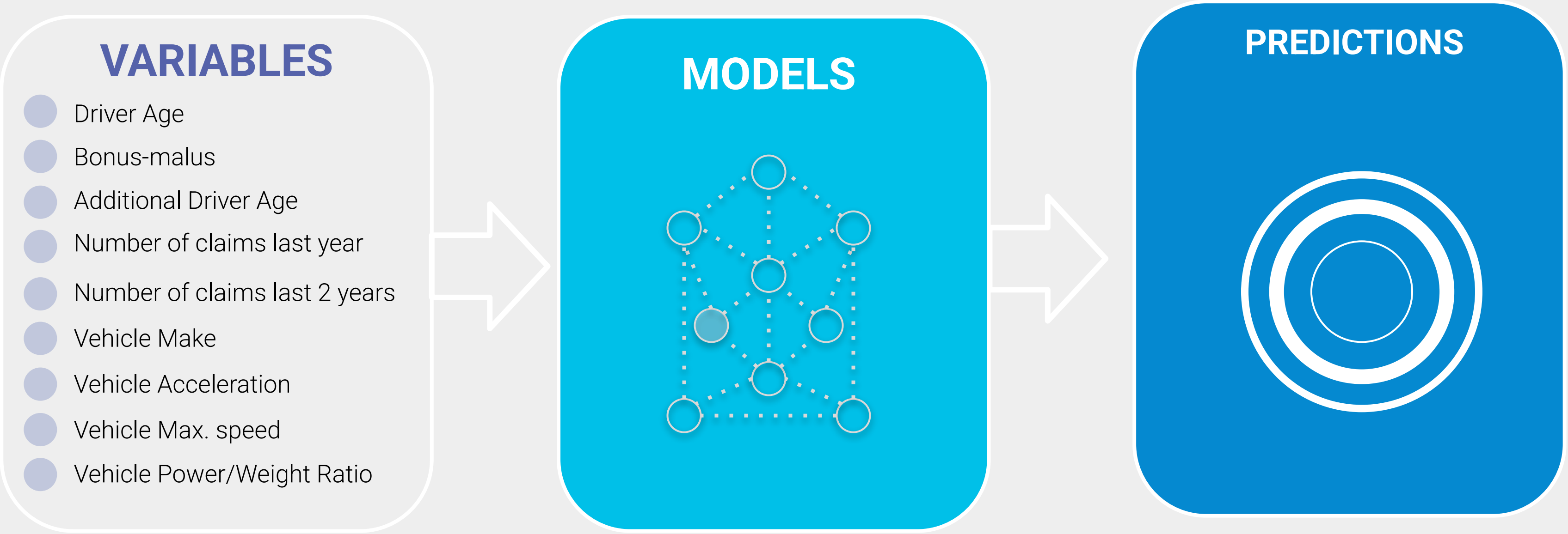
If I don't price properly, I'm exposed to Adverse Selection Risk



AKUR8

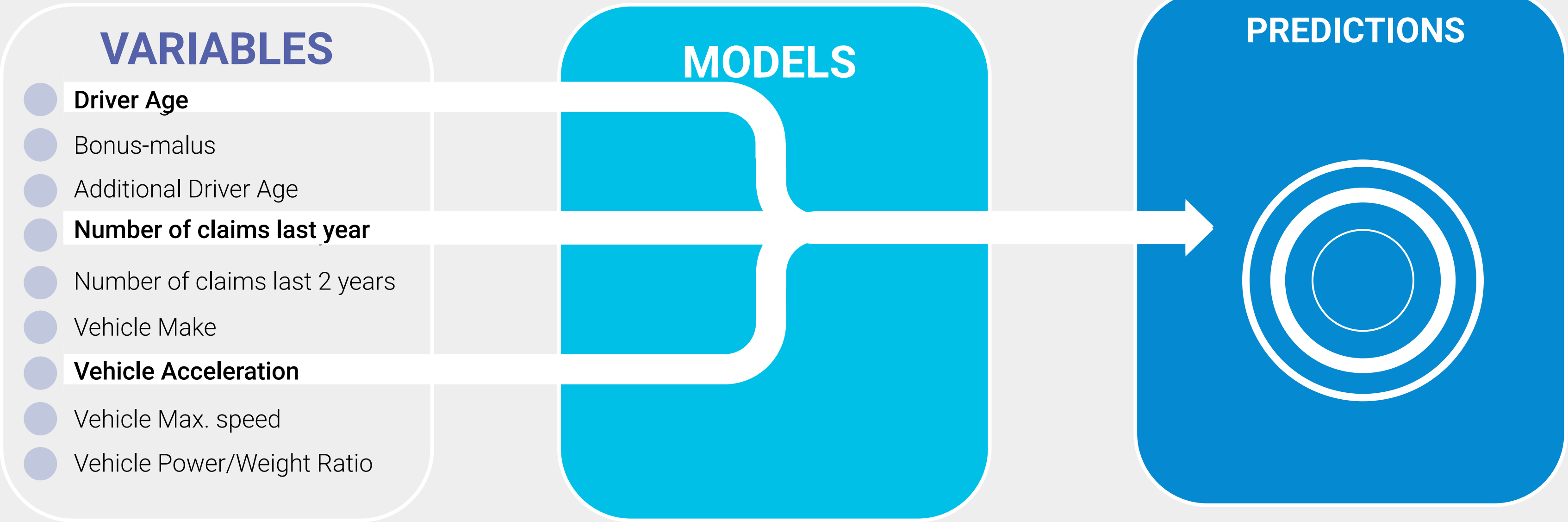
P&C Pricing: Different Approaches

Pricing classic approach



Pricing classic approach

- **Limited number** of effects impact the **predictions**
- Each effect has a clear & direct **impact** on the **predictions**

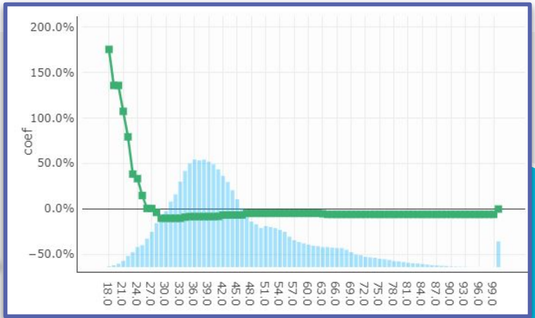


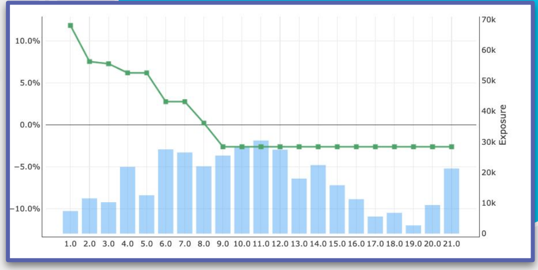
Possible to visualize, understand and modify the models

Pricing classic approach

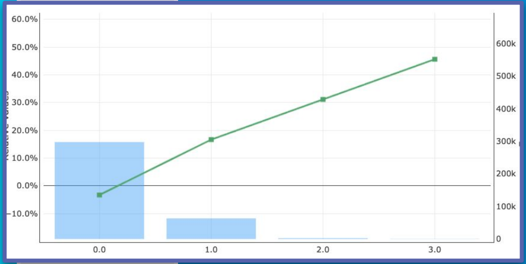
VARIABLES

- **Driver Age**
- Bonus-malus
- Additional Driver Age
- **Number of claims last year**
- Number of claims last 2 years
- Vehicle Make
- **Vehicle Acceleration**
- Vehicle Max. speed
- Vehicle Power/Weight Ratio





PREDICTIONS



Example: Rating Table

Number of Claims last year	
0	0.85
1	1.17
2+	1.40

Driver Age	
18	2.75
19	2.30
20	2.00
21	1.85
22	1.72
...	...

Vehicle Acceleration	
1.0	1.12
2.0	1.07
3.0	1.04
4.0	1.01
5.0	0.97
...	...

Base	$ 100

Estimation for a 20 years hold policyholder with no claims in the last year that drives a vehicle with an acceleration of 2.0

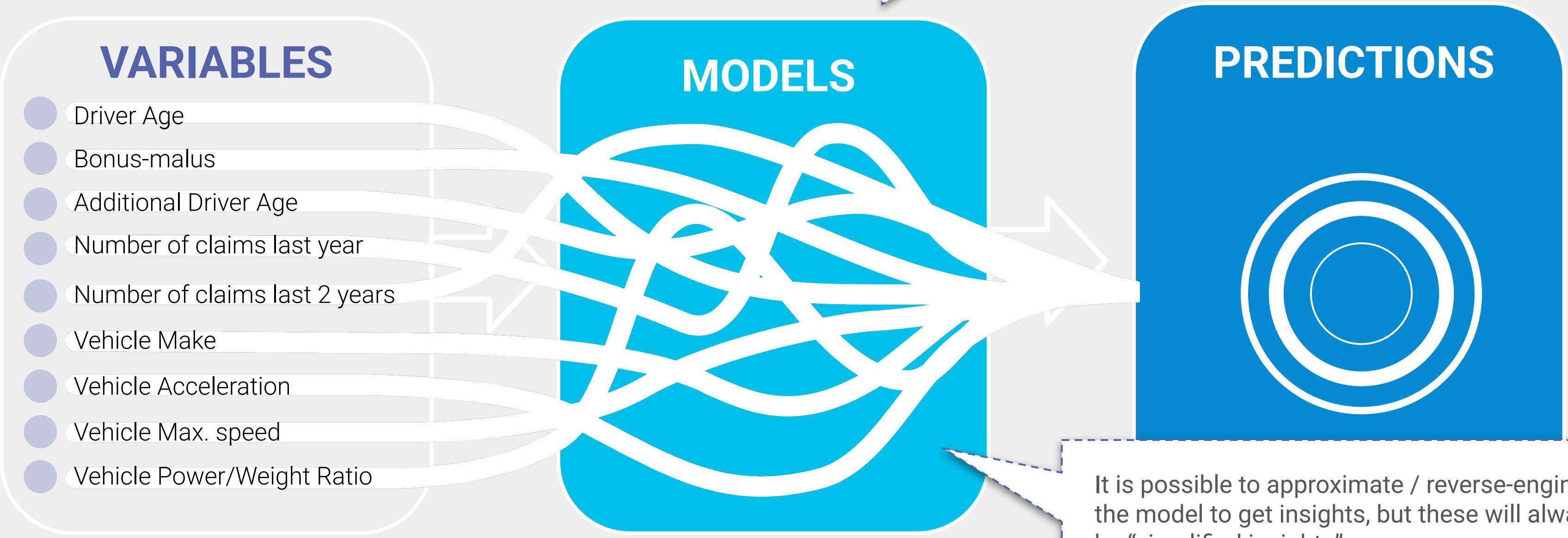
Prediction = $ 100 × 0.85 × 2.00 × 1.07
= $ 181.90

AKUR8

Black-Box Machine Learning

- No **selection** of variables used
- All variables effects depend of other variables

Not possible to understand the effects of the variables.



It is possible to approximate / reverse-engineer the model to get insights, but these will always be “simplified insights”.
Mismatch between what you see & what you put to production

Generalized Linear Models in Pricing

Generalized Linear Models

GLMs are extensively used for P&C Pricing due to their flexibility and interpretability

$$g(E(Y_i)) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip},$$

Target
- Frequency
- Severity
- Pure Premium

Coefficients (variable effects)

Estimation technique:
Maximum Likelihood

$$\widehat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta} \in \mathbb{R}^{p+1}} LLH(\boldsymbol{\beta}, \boldsymbol{y})$$

Variables
- Driver Age
- Vehicle Age
- Vehicle Brand
- etc…

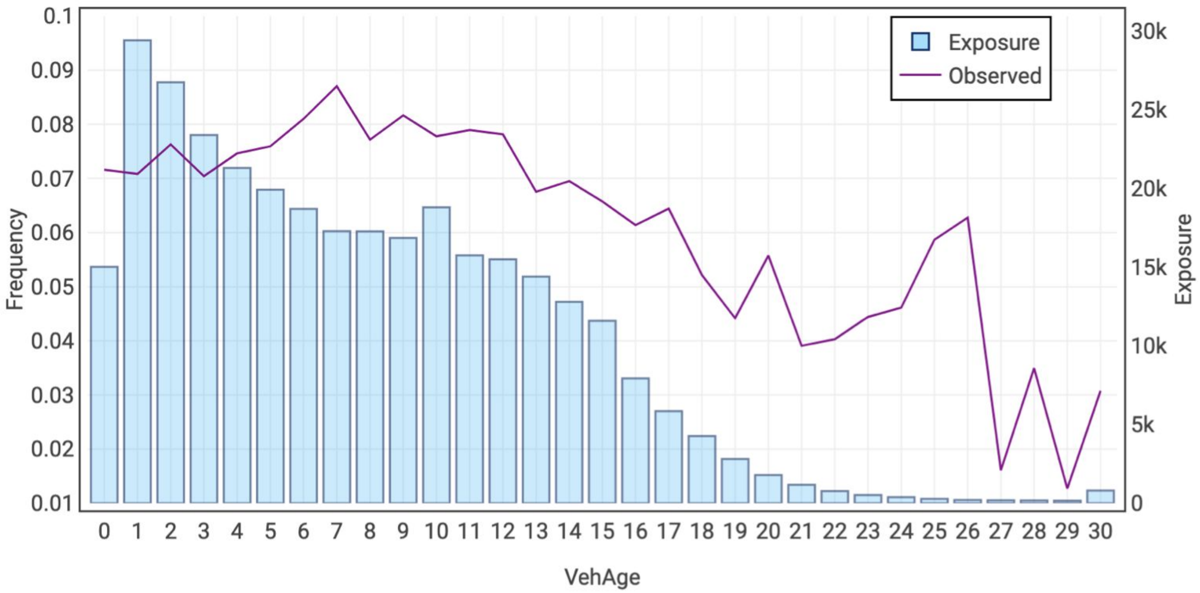
Pure Premium = Frequency × Severity

Challenges of GLM: Non-Linear Effects

Challenges of GLM: Non-Linear Effects

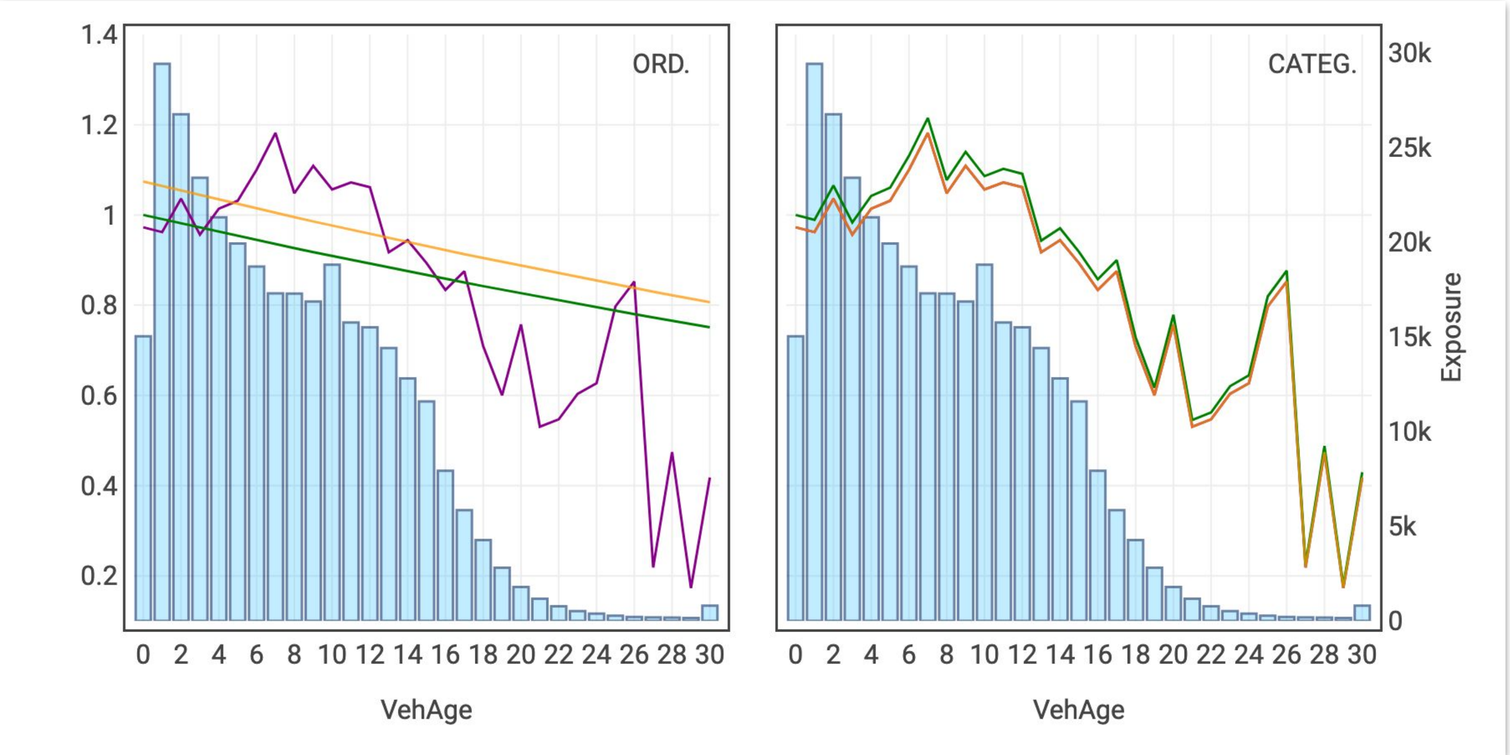
Modeling an ordinal variable

The Vehicle Age (VehAge) variable represents the age of the insured car (in years). The figure below provides the univariate effect of the variable.



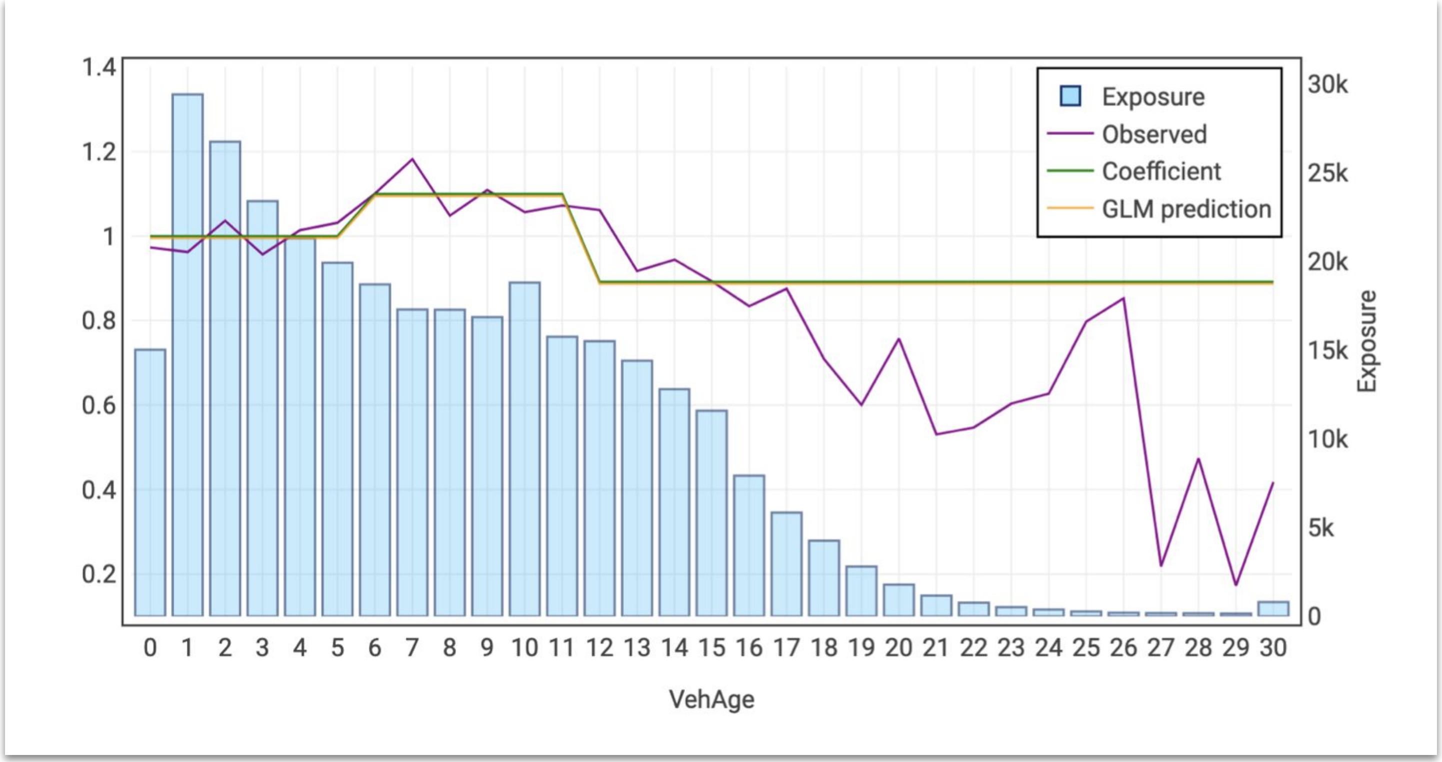
Challenges of GLM: Non-Linear Effects

Good GLMs require extensive feature engineering



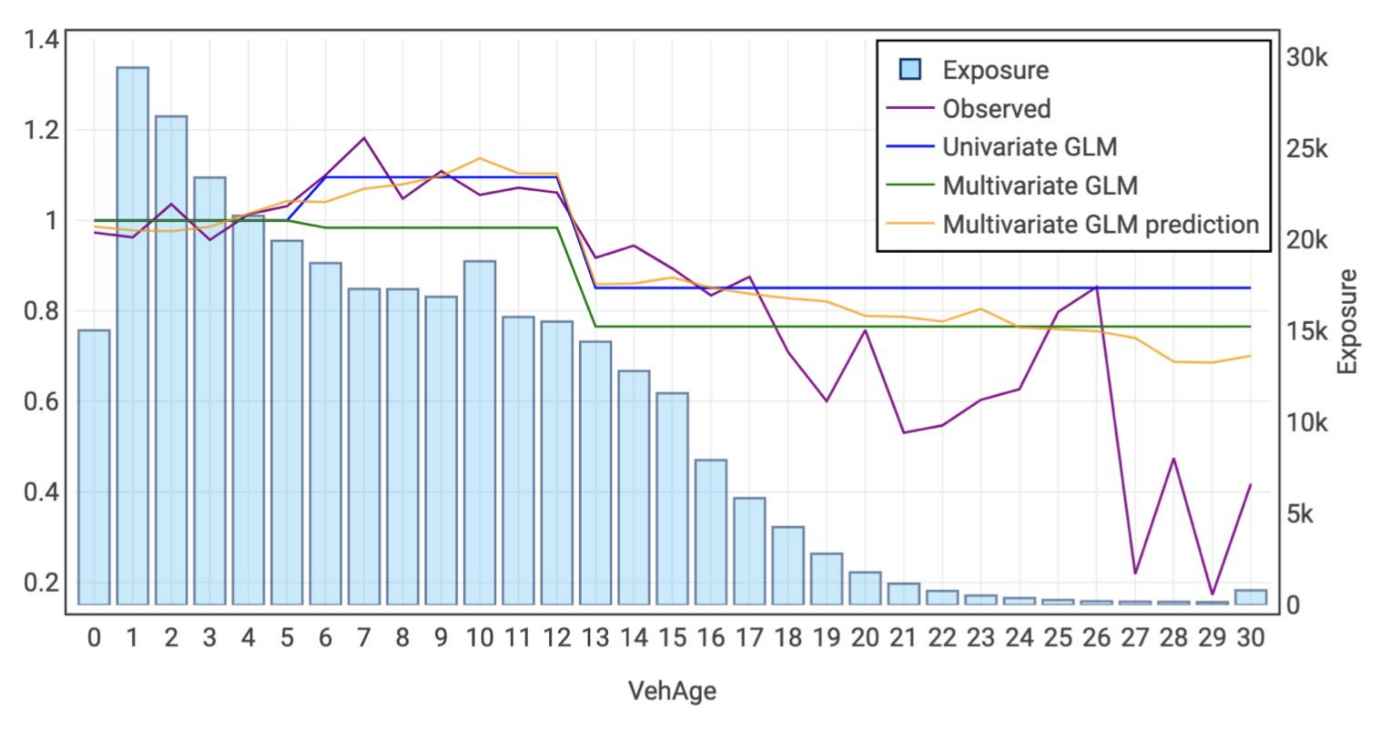
Challenges of GLM: Non-Linear Effects

Three splits - univariate fit



Challenges of GLM: Non-Linear Effects

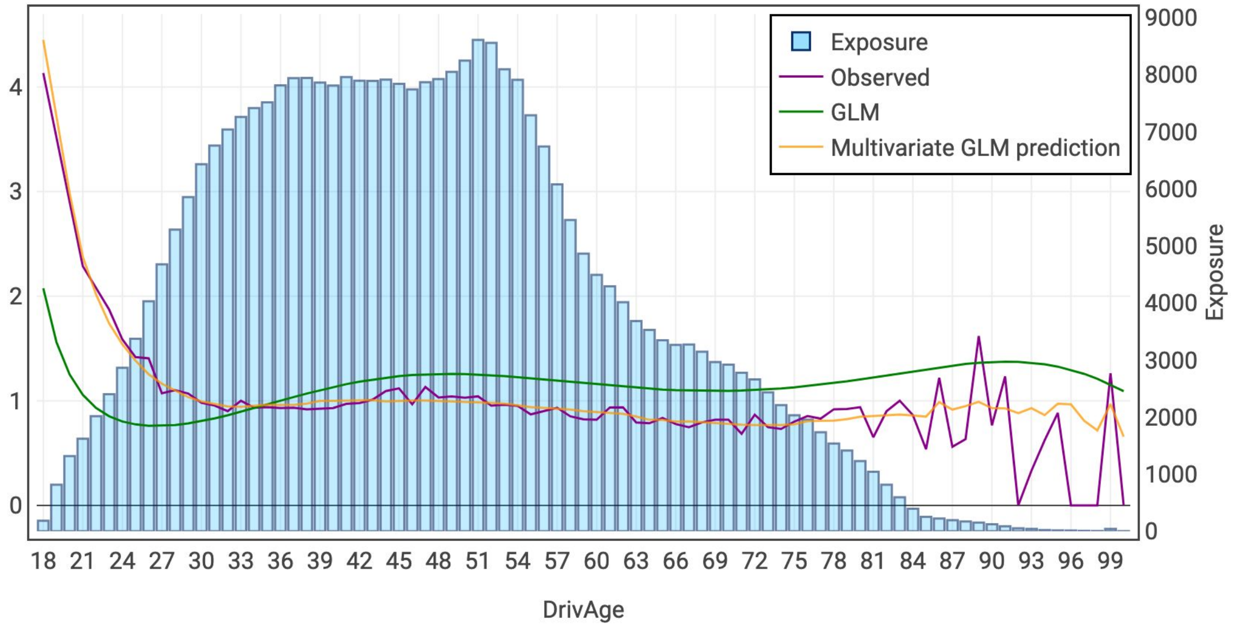
Three splits - multivariate



Challenges of GLM: Non-Linear Effects

Non-linear numeric variable fitting

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



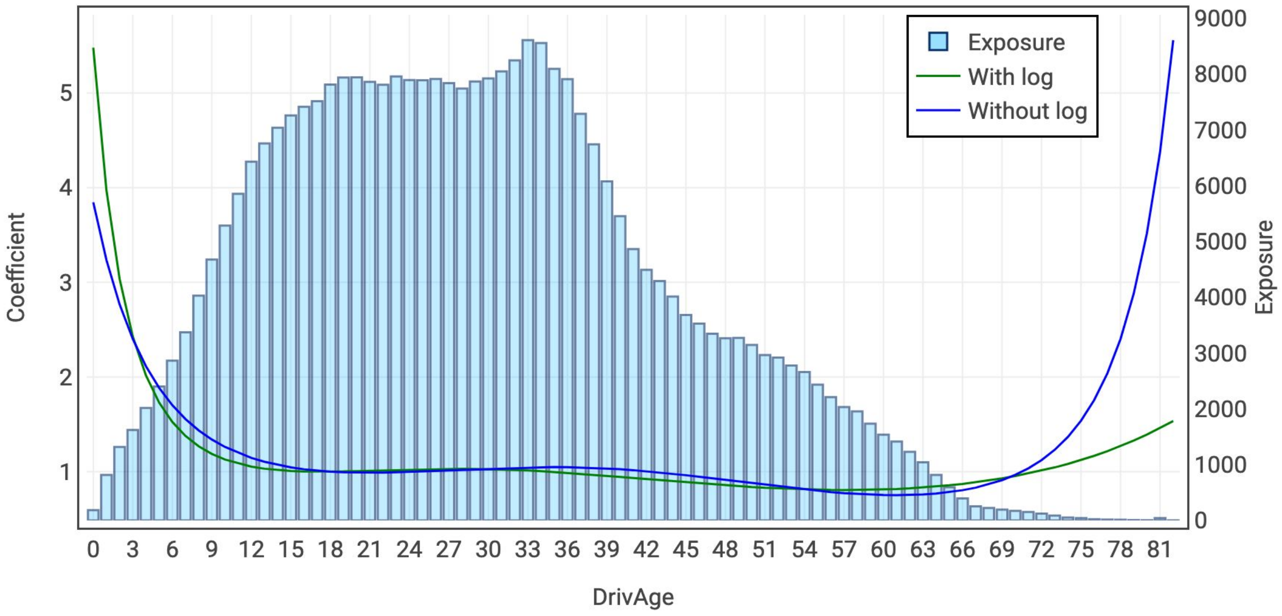
AKUR8

CONFIDENTIAL

Challenges of GLM: Non-Linear Effects

Non-linear numeric variable fitting and extrapolation

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



Challenges of GLM: Non-Linear Effects

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

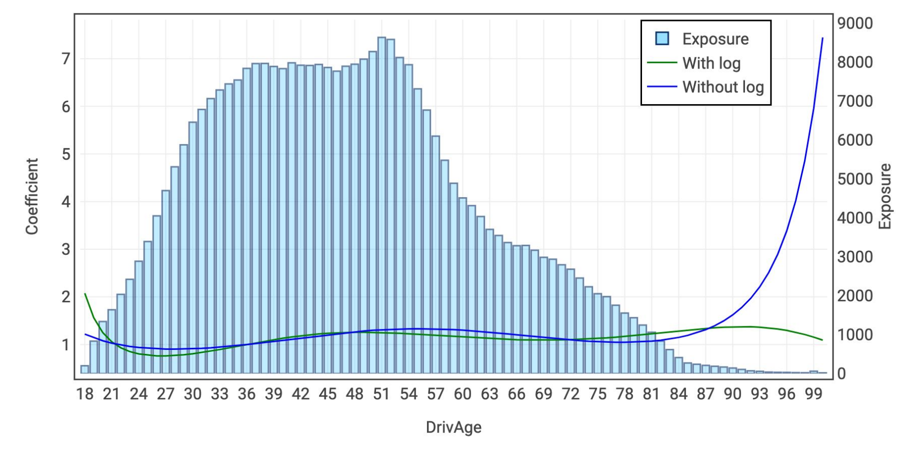
The estimates of the model are wildly different for the older tail - even if both models are statistically sound.

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



Challenges of GLM: Non-Linear Effects

Transformations take a lot of time and are arbitrary

One often spends weeks evaluating transformations in a multivariate model with multiple coverages/perils

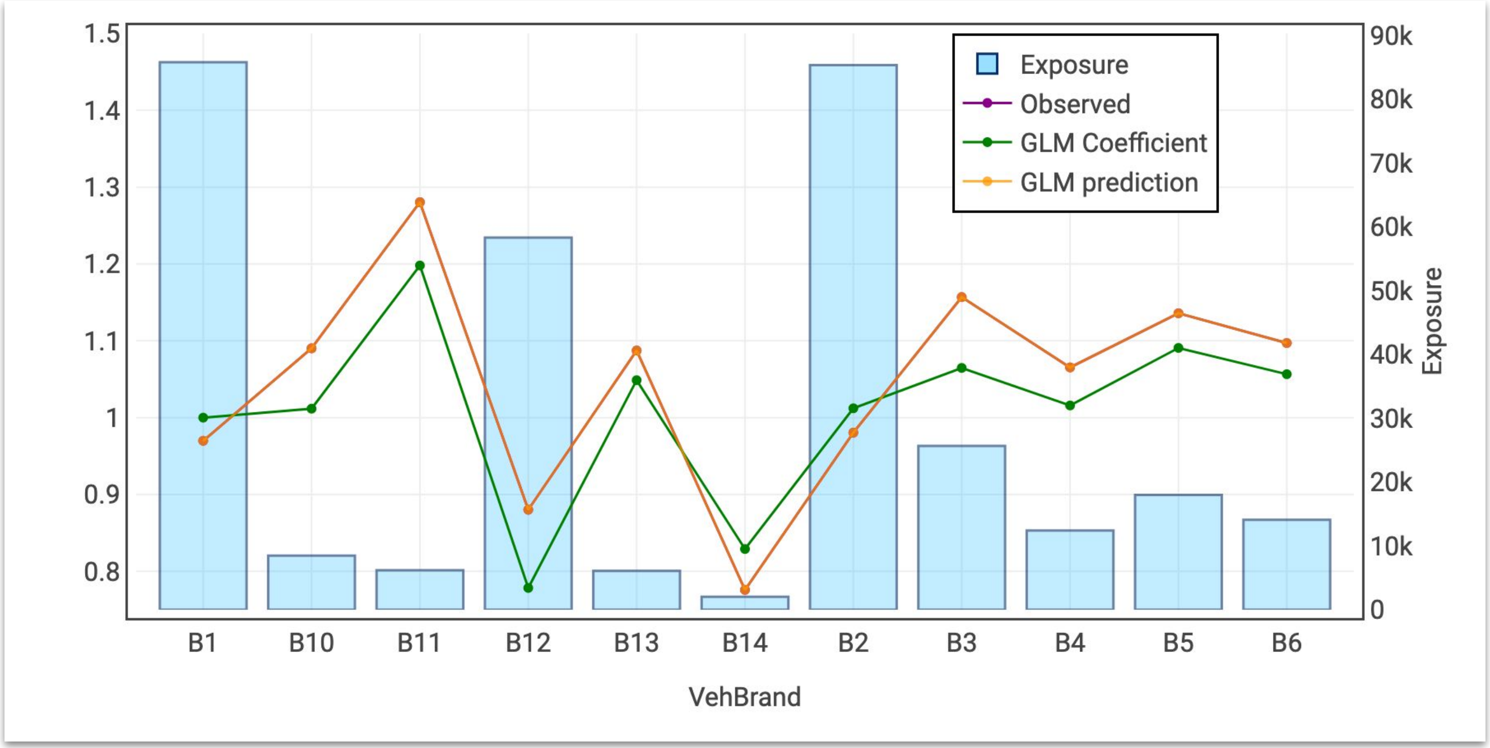
Every transformation is an opportunity to build a better model or undermine the model:

- Logging
- Square rooting
- Polynomial of varying degrees
- Reciprocal
- Capping/Flooring/Winsorizing
- Hinge/One-Hot Encoding
- Binning/Discretizing
- Splines (Linear/Cubic/B-Splines)
- Yeo-Johnson & Box-Cox
- Thermometer/Unary Encoding
- many many more…..

Challenges of GLM: Variable Selection

Challenges of GLM: Variable Selection

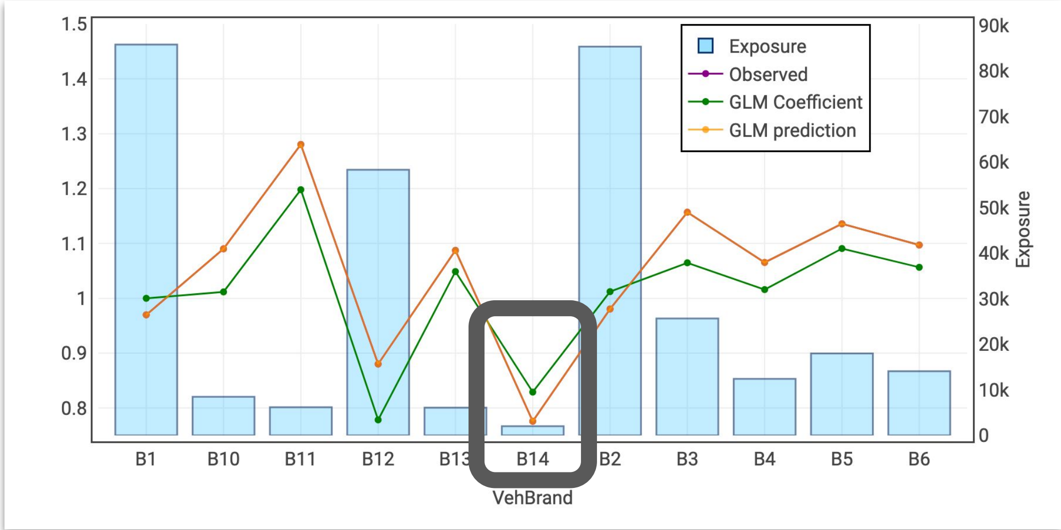
Categorical variables



Challenges of GLM: Variable Selection

Categorical variables

What do I do with B14? Include or exclude?
4.6% p-value but 20% discount

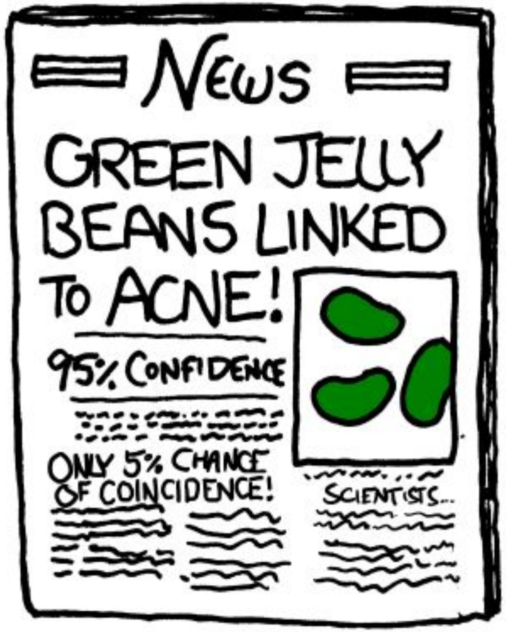


	GLM Coefficient	p-value
B1	1.00000	nan
B10	1.01167	0.78277
B11	1.19804	0.00005
B12	0.77841	0.00000
B13	1.04876	0.31601
B14	0.82906	0.04617
B2	1.01219	0.50507
B3	1.06474	0.01297
B4	1.01589	0.65058
B5	1.09073	0.00289
B6	1.05653	0.09113

Challenges of GLM: Variable Selection

P-Values are widely used but come with complications (source)

- Threshold is arbitrary
- Subject to p-hacking, data dredging
- Potential overfitting
- Not analogous to credibility
- Loses meaning with penalization
- Binary decision, ignores effect size
- Multiple comparison or multiplicity
- Instability across subsets/folds
- Depends on the base level



Takeaways on GLMs

GLMs are the de facto standard in Non-Life Pricing but they have pros and cons

Pros

- Flexible to work with P&C actuarial data (frequency, severity and pure premium)
- Transparent and easy to explain
- Work well for segments with enough exposure

Cons

- Require a lot of iterative feature engineering to address non-linearity
- Significance tests are ambiguous and arbitrary
- Instability in segments with limited exposure

Penalized Regression

Penalized Regression

I can add a penalization term to the Maximum Likelihood estimator to prevent overfitting

GLM Regression

$$\widehat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta} \in \mathbb{R}^{p+1}} LLH(\boldsymbol{\beta}, \boldsymbol{y})$$

Maximum Likelihood GLM can bring to **overfitting**

Penalized Regression

$$\widehat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta} \in \mathbb{R}^{p+1}} \{LLH(\boldsymbol{\beta}, \boldsymbol{y}) - \underline{\lambda\, \text{Penalization}(\boldsymbol{\beta})}\} \qquad \lambda \geq 0$$

Ridge Regression

$$\widehat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta} \in \mathbb{R}^{p+1}} \left\{LLH(\boldsymbol{\beta}, \boldsymbol{y}) - \underline{\lambda \textstyle\sum_{j=1}^{p} \beta_j^2}\right\}$$

LASSO Regression

$$\widehat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta} \in \mathbb{R}^{p+1}} \left\{LLH(\boldsymbol{\beta}, \boldsymbol{y}) - \underline{\lambda \textstyle\sum_{j=1}^{p} |\beta_j|}\right\}$$

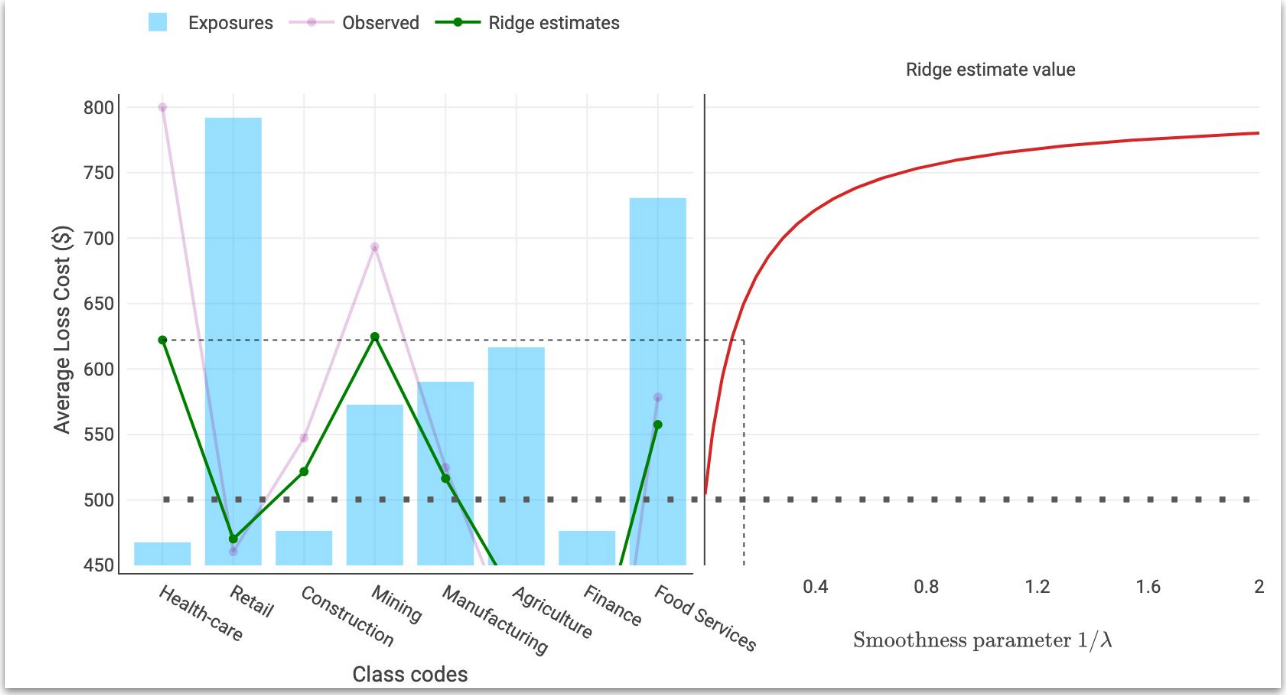
Ridge Regression

Ridge Regression

Impact of smoothness to Ridge estimates - Large λ

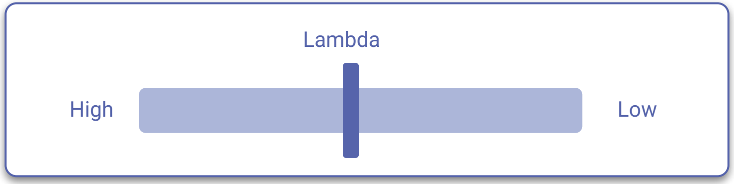
Large λ
(large penalty)

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



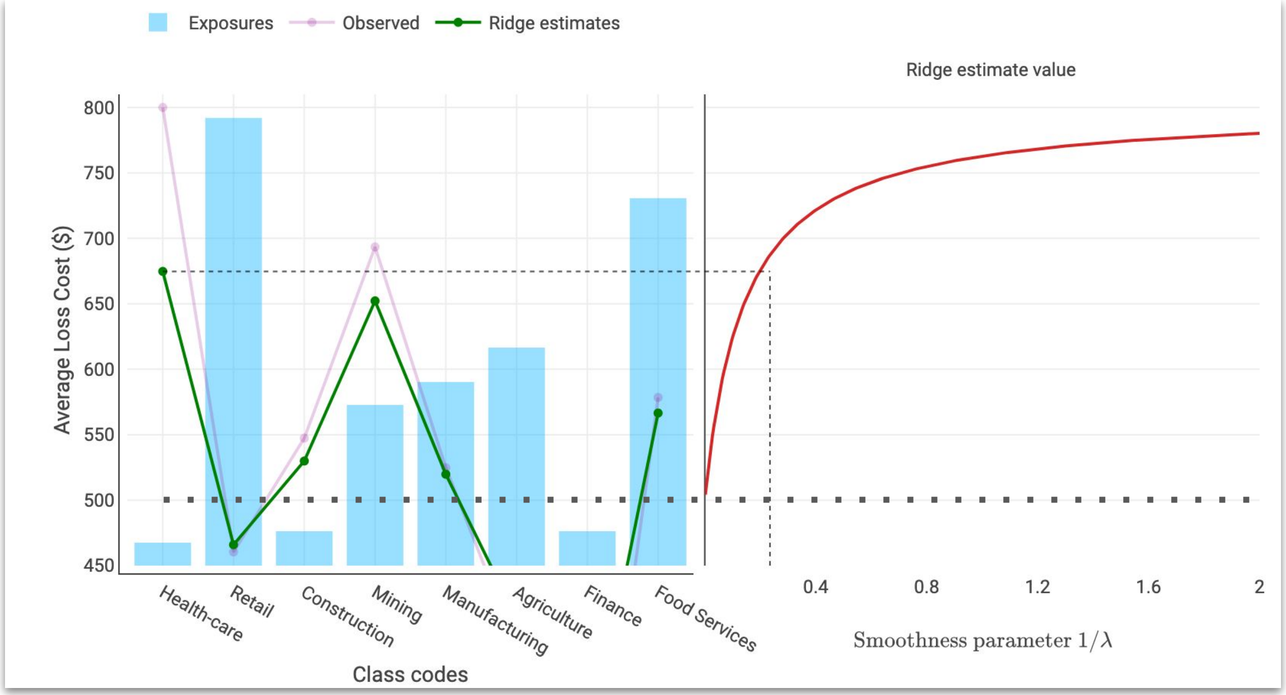
Ridge Regression

Impact of smoothness to Ridge estimates - Medium λ



Medium λ
(medium penalty)

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

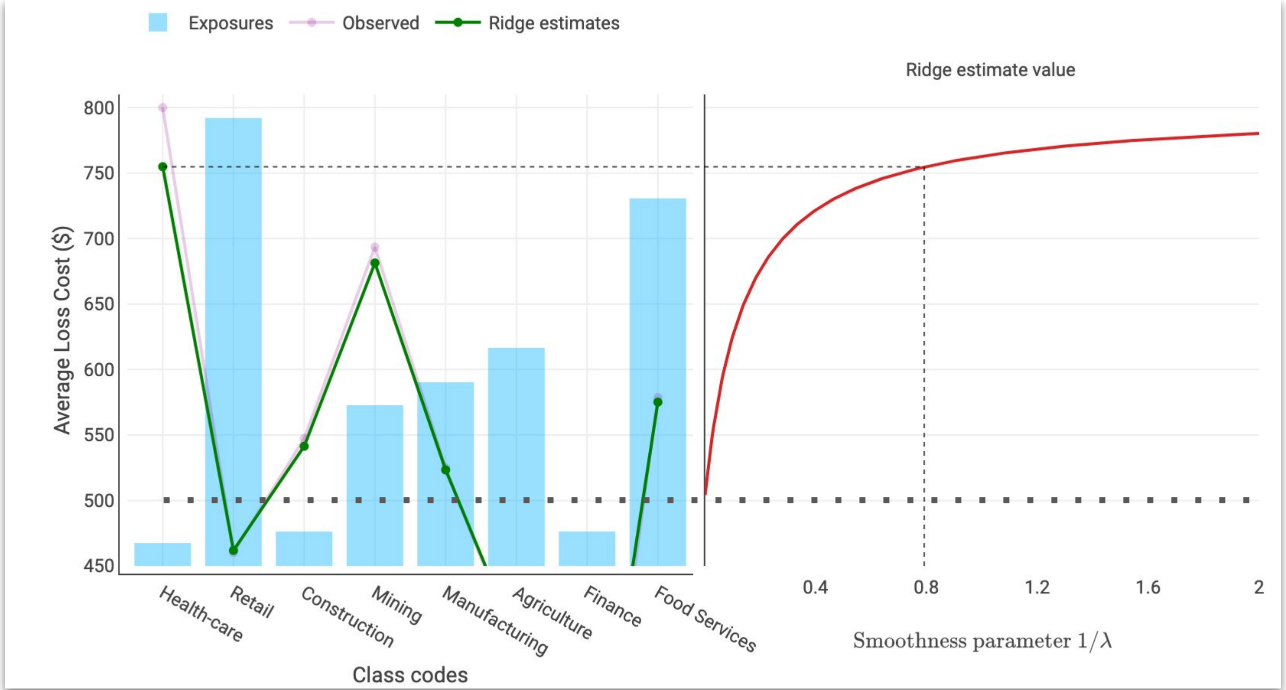


Ridge Regression

Impact of smoothness to Ridge estimates - Small λ

Small λ
(small penalty)

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



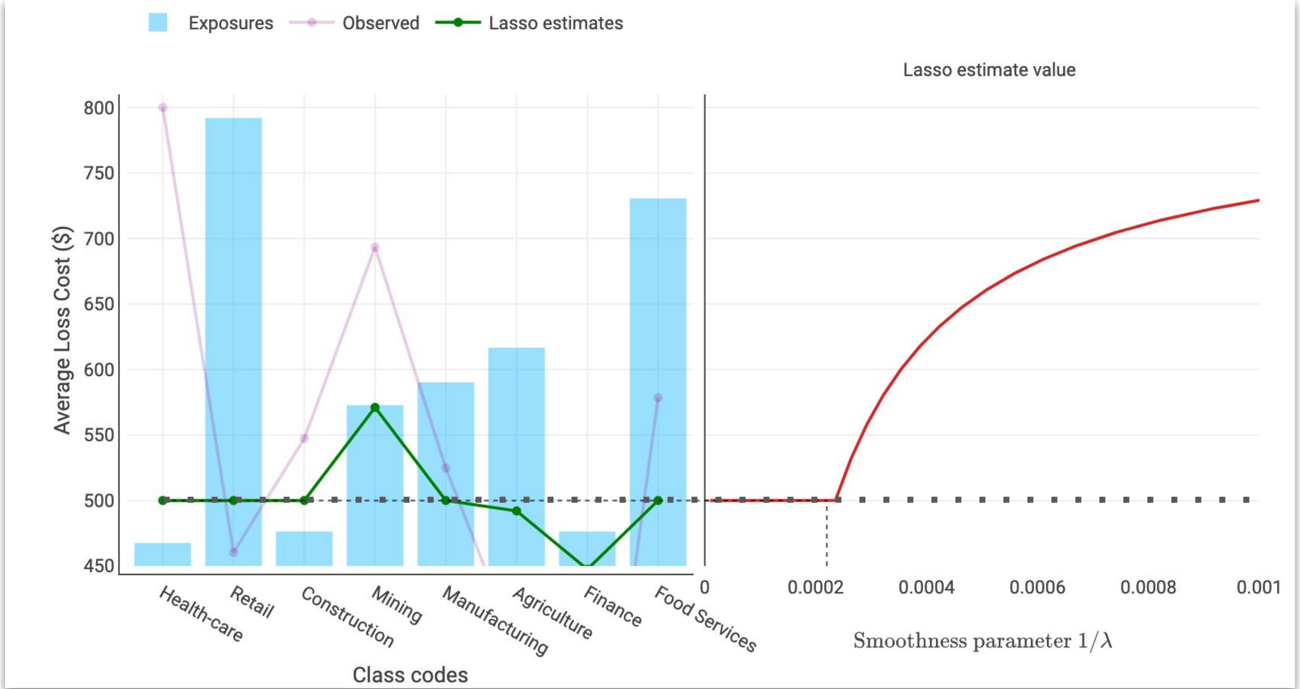
LASSO Regression

LASSO Regression

Impact of smoothness to LASSO estimates - Large λ

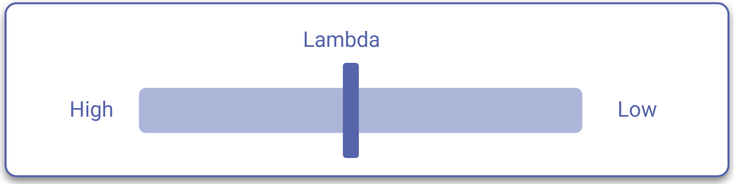
Large λ
(large penalty)

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



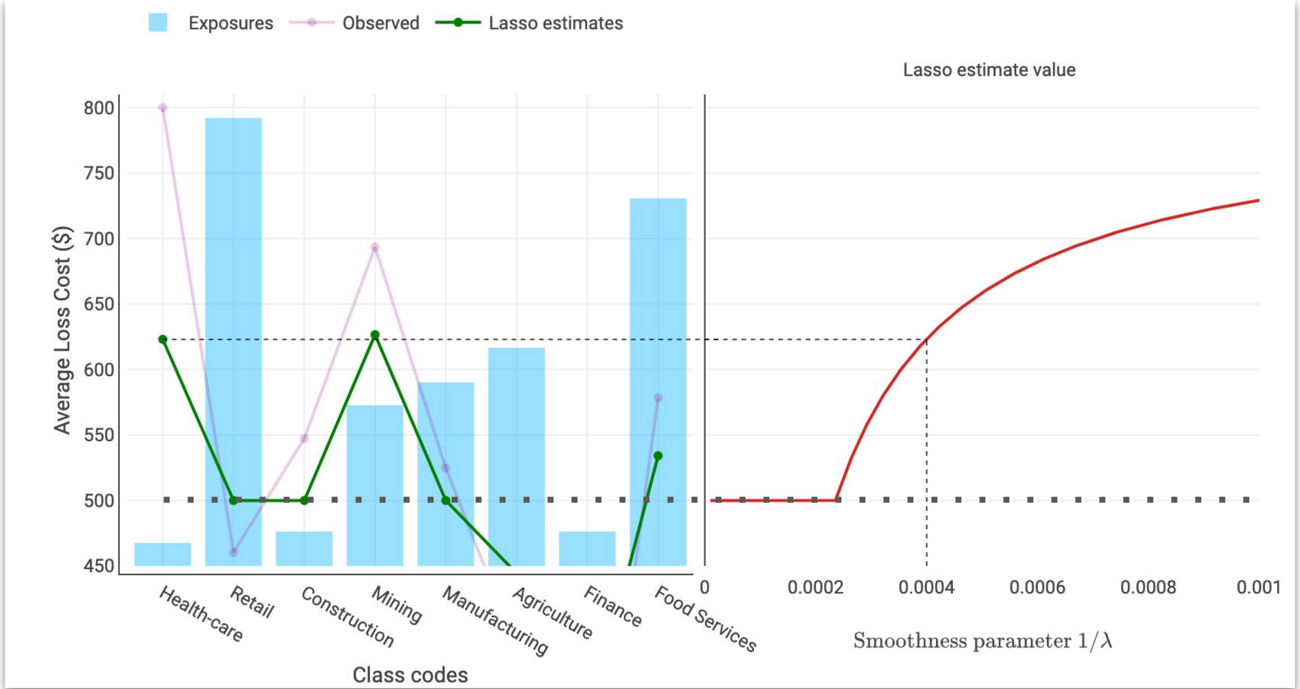
LASSO Regression

Impact of smoothness to LASSO estimates - Medium λ



Medium λ
(medium penalty)

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

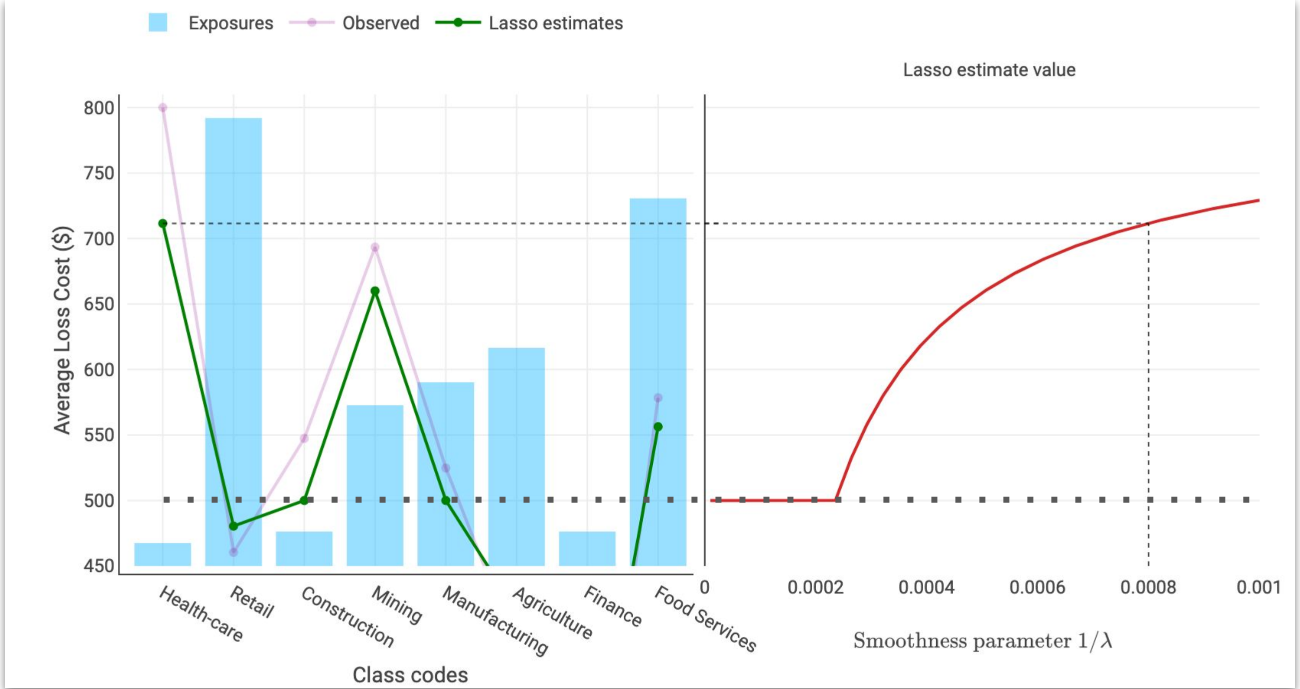


LASSO Regression

Impact of smoothness to LASSO estimates - Small λ

Small λ
(small penalty)

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

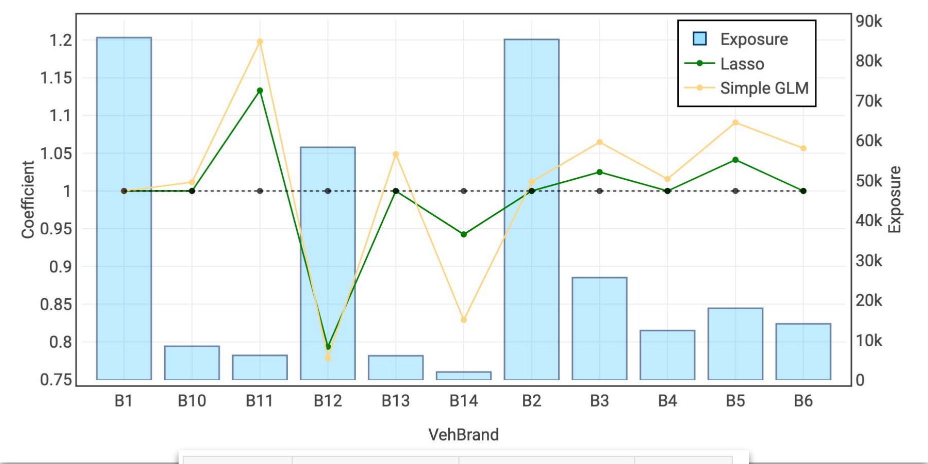


LASSO Regression

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.

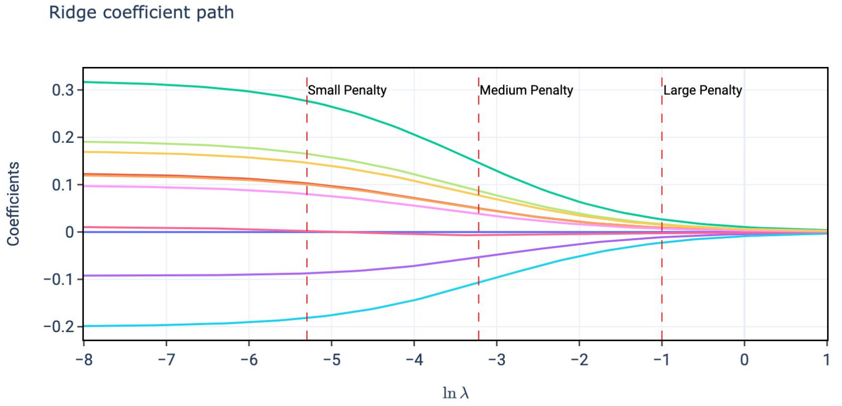


VehBrand	GLM Coefficients	Lasso Coefficients	P-Values
B1	1.00000	1.00000	nan
B10	1.01167	1.00000	0.78277
B11	1.19804	1.13311	0.00005
B12	0.77841	0.79358	0.00000
B13	1.04876	1.00000	0.31601
B14	0.82906	0.94246	0.04617
B2	1.01219	1.00000	0.50507
B3	1.06474	1.02499	0.01297
B4	1.01589	1.00000	0.65058
B5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

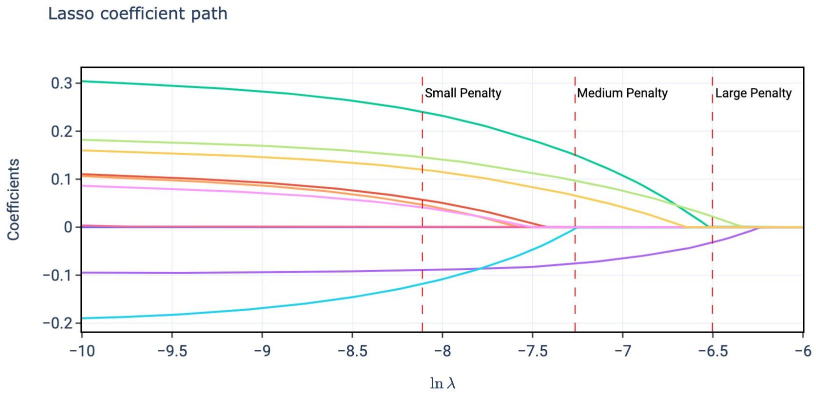
LASSO Regression can simplify model review by applying credibility considerations as well as automating coefficient removal.

Lasso and Ridge coefficients as functions of lambda

Ridge Regression

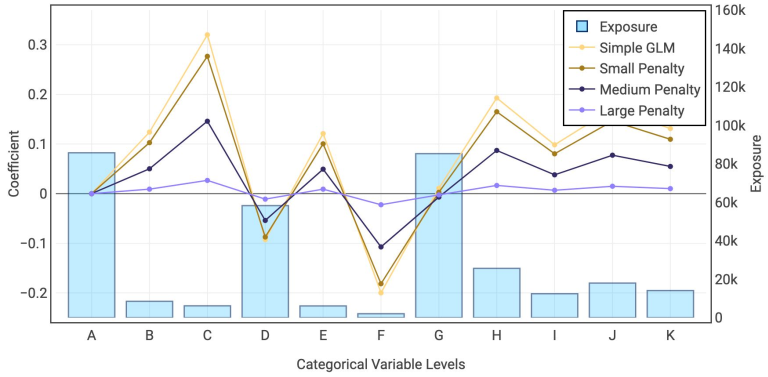


Lasso Regression

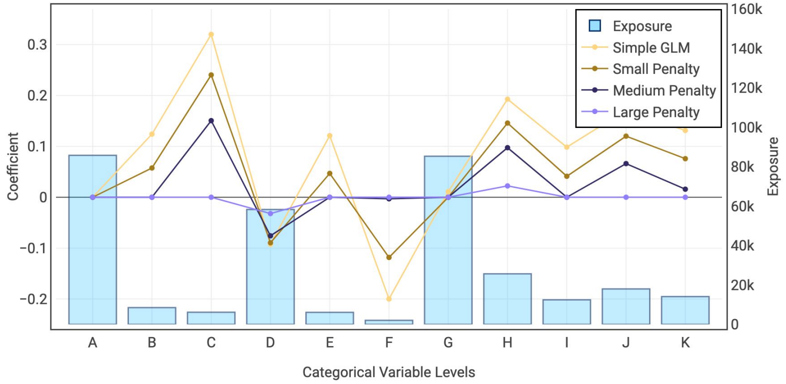


Coefficient graph for different values of lambda

Ridge Regression



Lasso Regression



Penalized Regression, Credibility and Bayesian Statistics

Quick Reminder… What is Credibility

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.

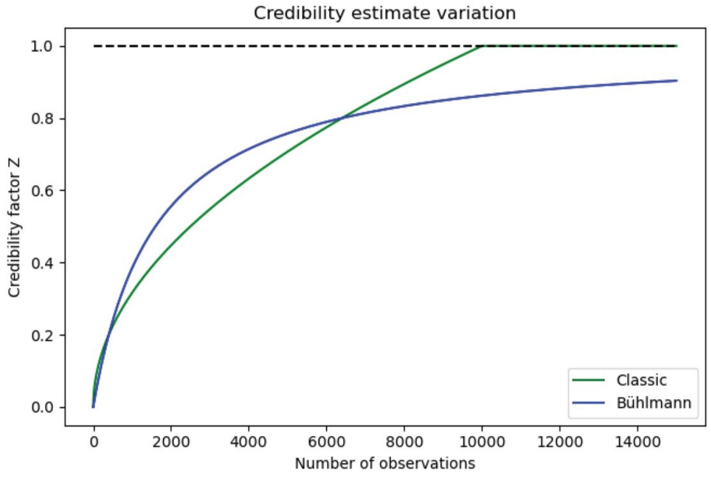
Credibility formula

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

Credibility factor

$0 \leq Z \leq 1$

Figure 2.1. Evolution of the credibility factor *Z* for a given estimate *j* as a function of the number of observations *n*. The *Z* for the classical credibility is computed using Equation 2.1 with *N* = 15,000. Bühlmann credibility uses *k* = 1,600 in Equation 2.2.



Penalized Regression, Credibility and Bayesian Statistics

It turns out that Ridge Regression is a Bayesian Estimator with Normal Prior

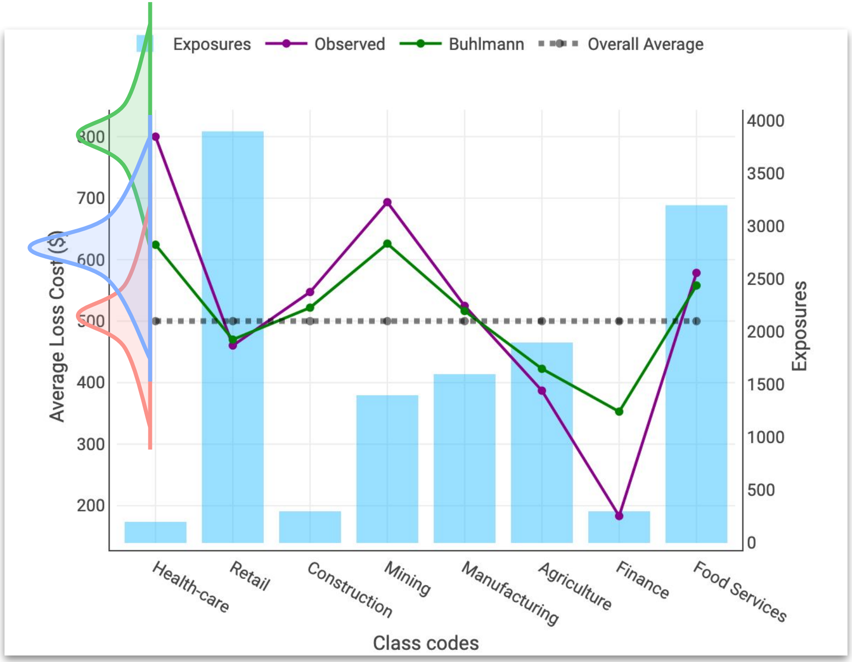
Ridge Regression

$$\widehat{\boldsymbol{\beta}}^{Ridge} = \arg\max_{\boldsymbol{\beta}\in\mathbb{R}^{p+1}}\left\{LLH(\boldsymbol{\beta},\boldsymbol{y}) - \lambda\sum_{j=1}^{p}\beta_j^2\right\}$$

Bayesian Estimator with Normal Prior

$$\beta_j \sim Normal\left(0,\frac{1}{\lambda}\right) \qquad \pi(\beta_j) \propto e^{-\lambda\beta_j^2}$$

$$\widehat{\beta_j}^{MAP} = \widehat{\beta_j}^{Ridge}$$



Penalized Regression, Credibility and Bayesian Statistics

It turns out that LASSO Regression is a Bayesian Estimator with Laplace Prior

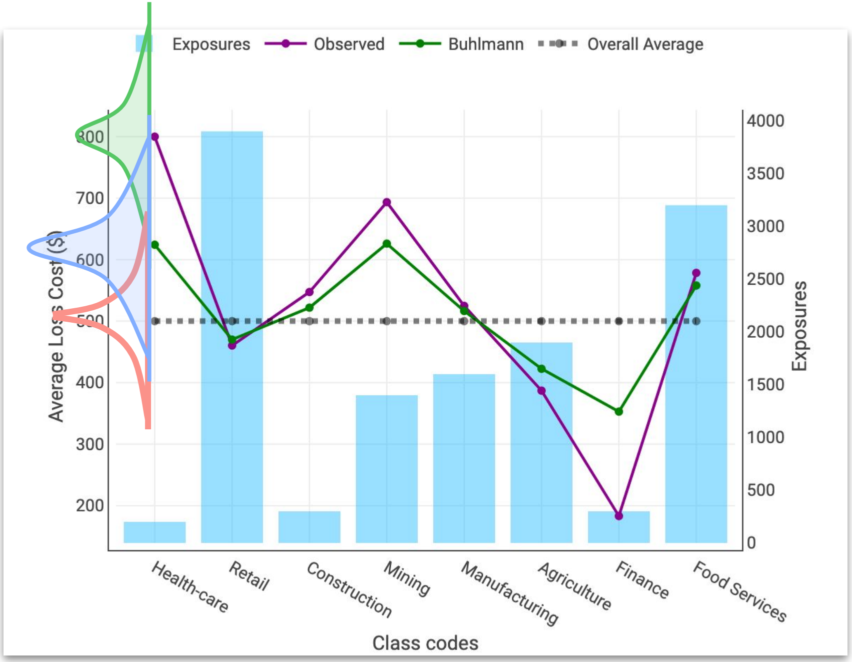
LASSO Regression

$$\widehat{\boldsymbol{\beta}}^{LASSO} = \arg\max_{\boldsymbol{\beta}\in\mathbb{R}^{p+1}}\left\{LLH(\boldsymbol{\beta},\boldsymbol{y}) - \lambda\textstyle\sum_{j=1}^{p}|\beta_j|\right\}$$

Bayesian Estimator with Laplace Prior

$$\beta_j \sim Laplace\left(0,\frac{1}{\lambda}\right) \qquad \pi(\beta_j) \propto e^{-\lambda|\beta_j|}$$

$$\widehat{\beta_j}^{MAP} = \widehat{\beta_j}^{LASSO}$$



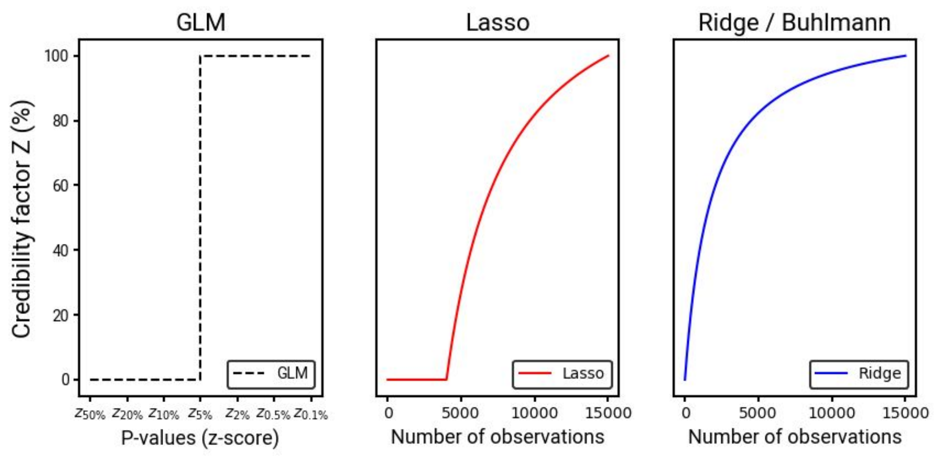
Penalized Regression, Credibility and Bayesian Statistics

Ridge, LASSO and p-value can be seen as credibility methods

p-value selection can be seen as a credibility approach that gives either **0% or 100% credibility**.

Ridge and LASSO introduce a **gradient of credibility**.

LASSO produces a feature selection that favors **sparsity**, or **model simplicity**.



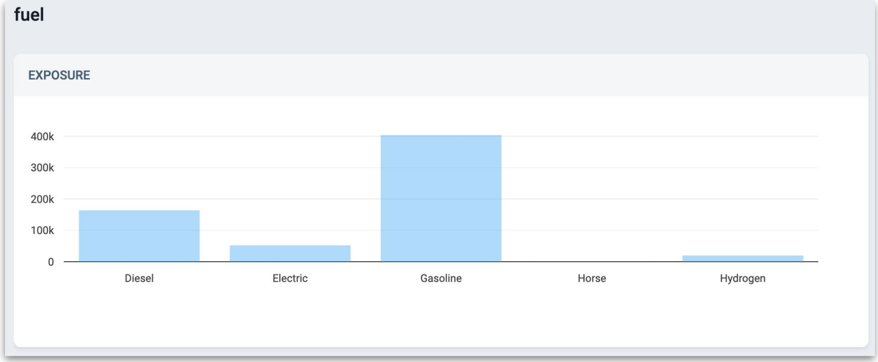
Non-credible coefficients are automatically removed when using a LASSO penalty parameter. The selection of an appropriate LASSO penalty parameter replaces p-value significance testing.

From Classical LASSO to Derivative LASSO

From Classical LASSO to Derivative LASSO

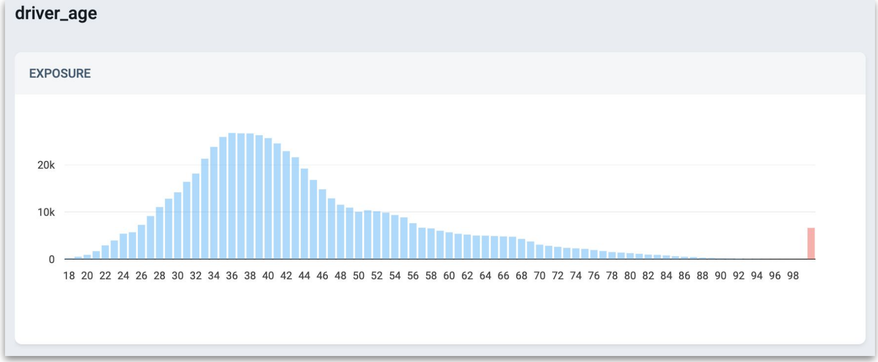
Variable Inputs are Ordinal or Categorical

Categorical variables have no intuitive link between levels. Each category is distinct.



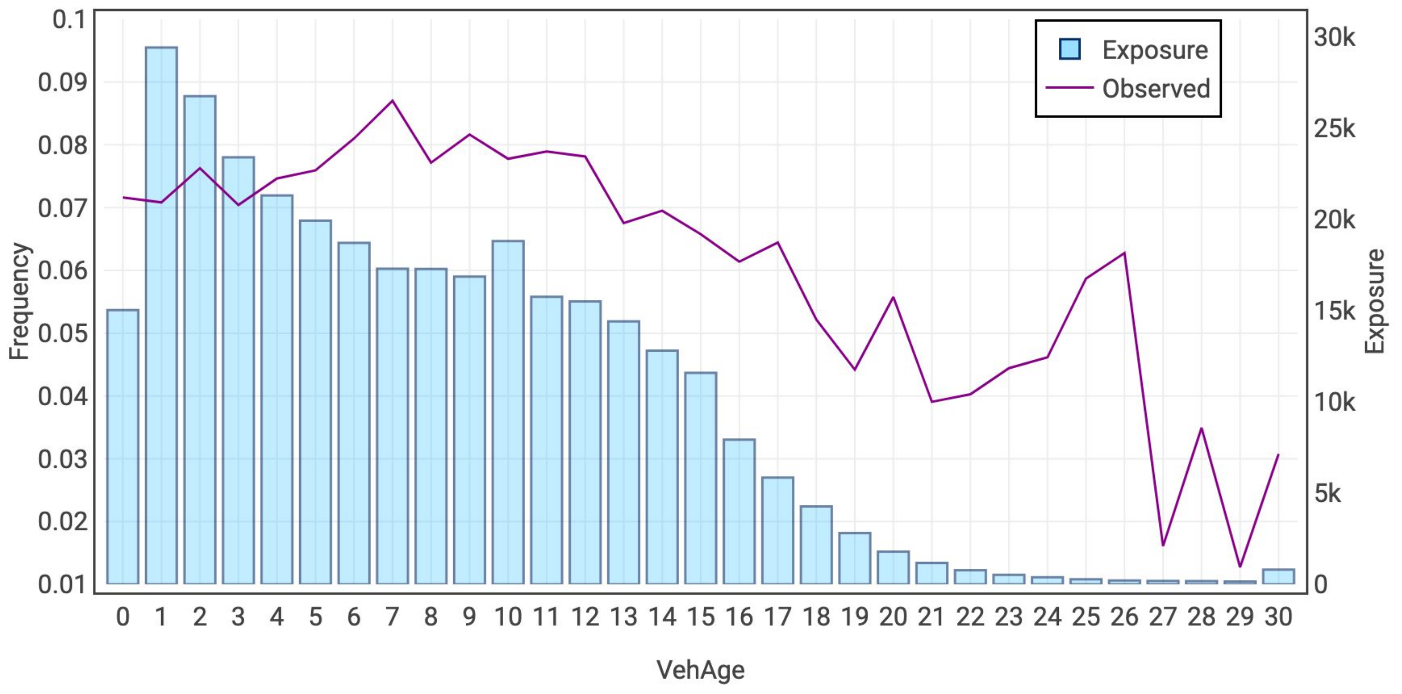
Ordinal variables have an intuitive link between adjacent levels. Each level is linked to its upper and lower neighbors.

Could also be A, B, C, D, E….



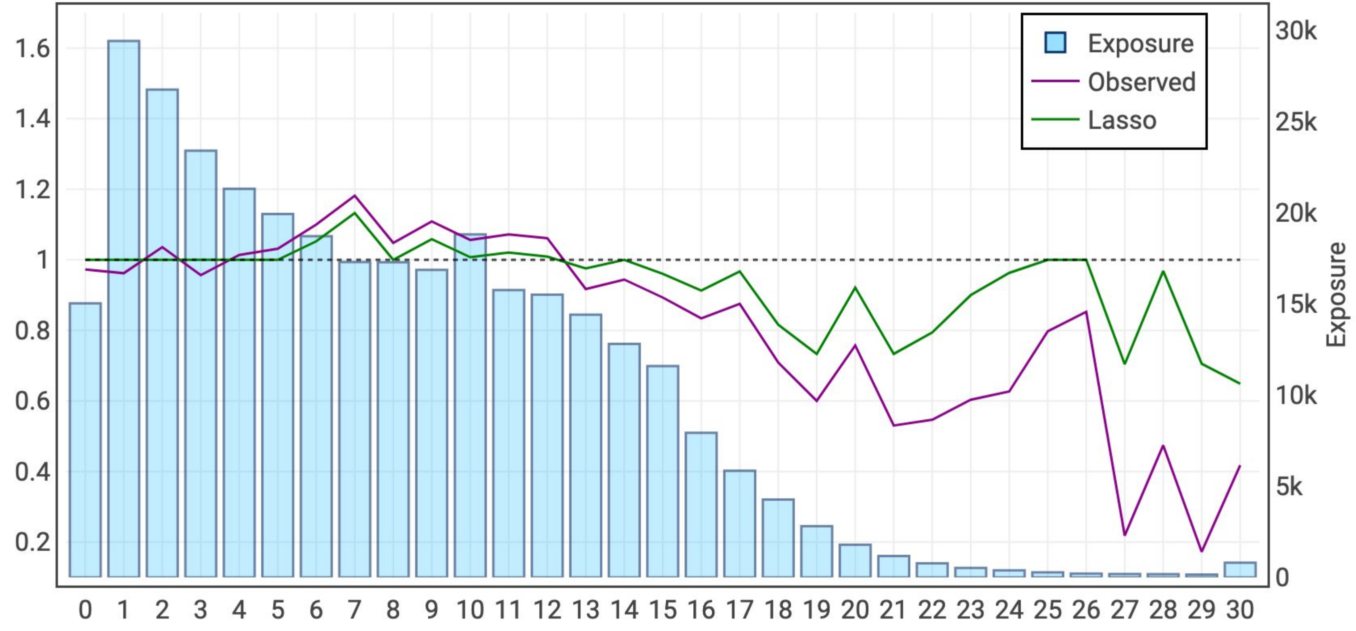
From Classical LASSO to Derivative LASSO

Ordinal variable example: Vehicle age



From Classical LASSO to Derivative LASSO

Classical LASSO penalization is not adapted for ordinal variables

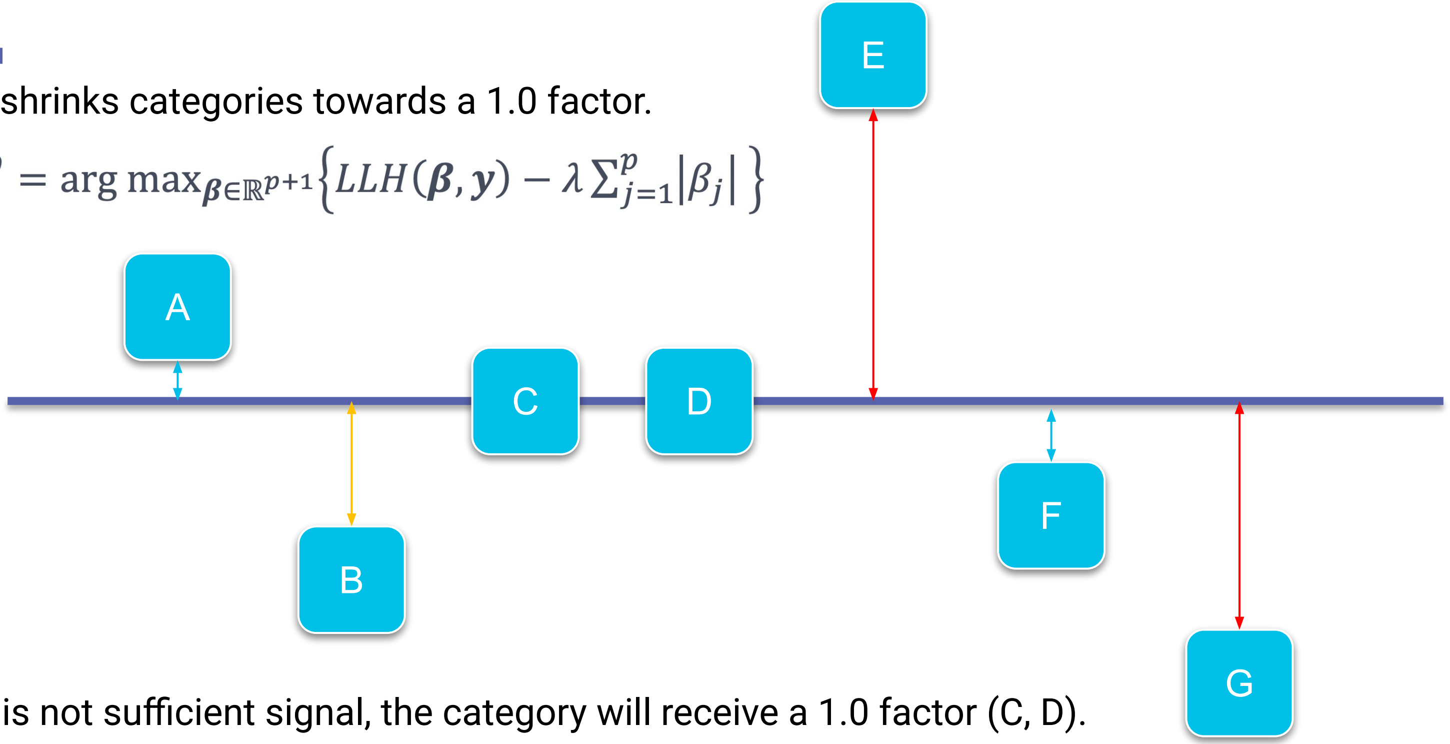


From Classical LASSO to Derivative LASSO

Impact of Classical LASSO penalty: it shrinks towards 1.0

LASSO shrinks categories towards a 1.0 factor.

$$\widehat{\boldsymbol{\beta}}^{LASSO} = \arg\max_{\boldsymbol{\beta}\in\mathbb{R}^{p+1}}\left\{LLH(\boldsymbol{\beta},\boldsymbol{y}) - \lambda \sum_{j=1}^{p}|\beta_j| \right\}$$



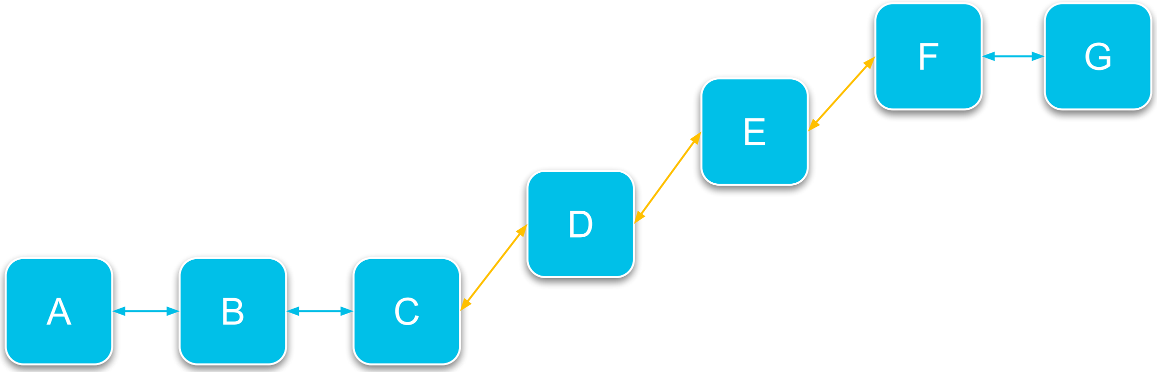
If there is not sufficient signal, the category will receive a 1.0 factor (C, D).

From Classical LASSO to Derivative LASSO

Impact of Derivative LASSO penalty: it shrinks variations

The **derivative** LASSO connects categories between their neighbors

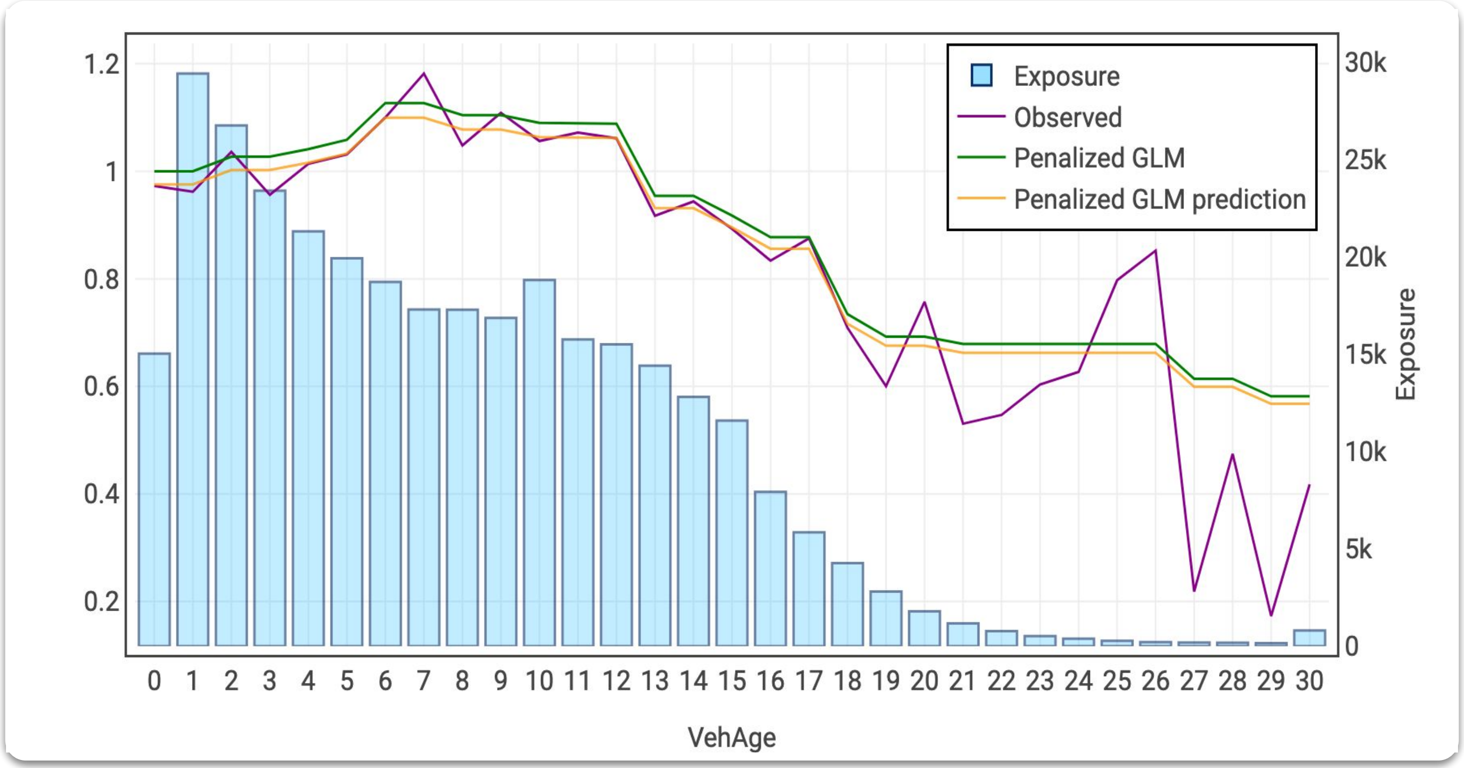
$$\widehat{\boldsymbol{\beta}}^{DL} = \arg\max_{\boldsymbol{\beta} \in \mathbb{R}^{p+1}} \left\{ LLH(\boldsymbol{\beta}, \boldsymbol{y}) - \lambda \sum_{j=1}^{p-1} \left| \beta_{j+1} - \beta_j \right| \right\}$$



If there is not sufficient signal, the category will receive the coefficient of the previous level (B, C, G).

Derivative LASSO

Derivative LASSO detects and incorporates non-linearities



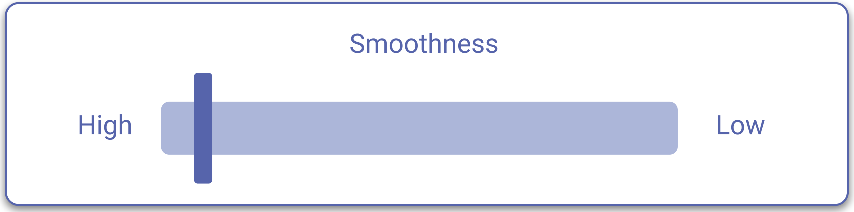
Derivative LASSO

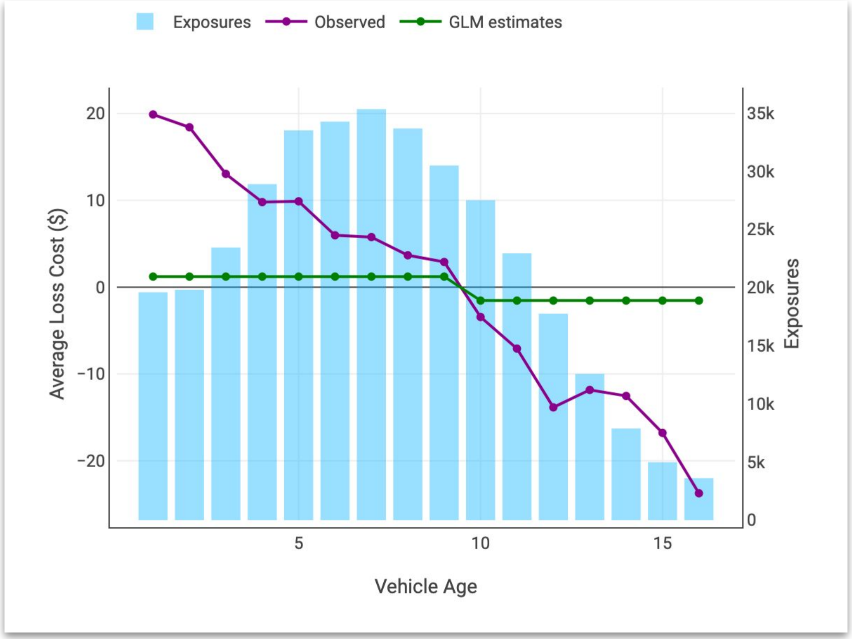
Influence of the smoothness

Derivative lasso **natively incorporates non-linear effects**.

To control the robustness and credibility, derivative lasso requires the selection of a **single credibility-based parameter**:

the **smoothness**





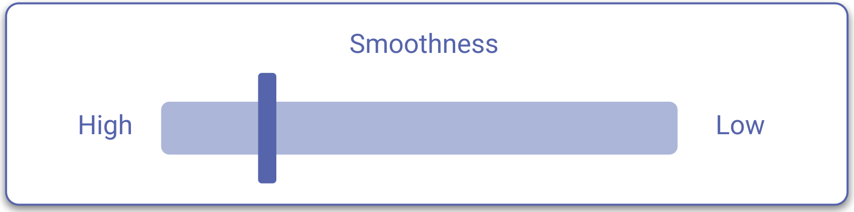
Derivative LASSO

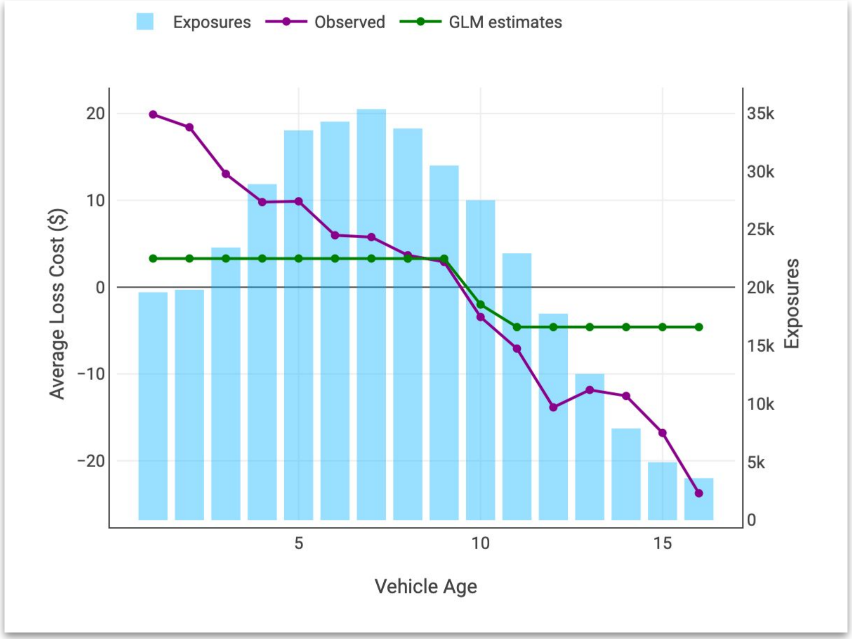
Influence of the smoothness

Derivative lasso **natively incorporates non-linear effects**.

To control the robustness and credibility, derivative lasso requires the selection of a **single credibility-based parameter**:

the **smoothness**





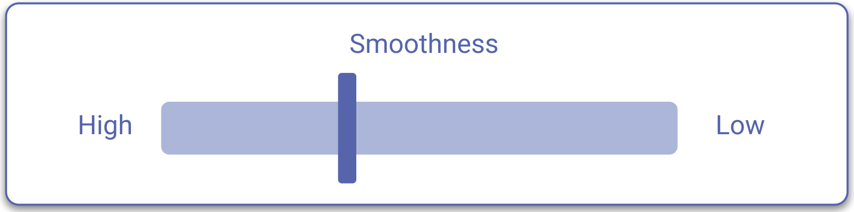
Derivative LASSO

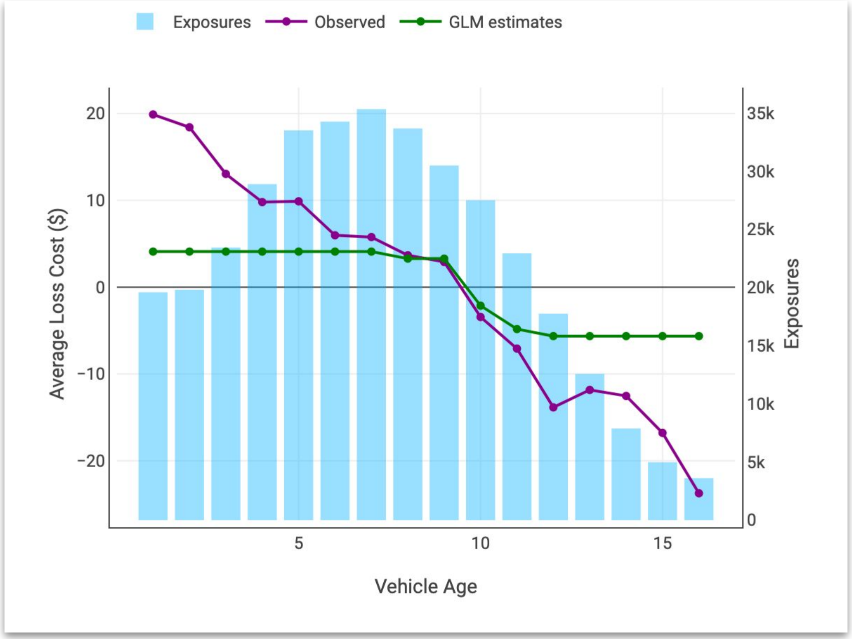
Influence of the smoothness

Derivative lasso **natively incorporates non-linear effects**.

To control the robustness and credibility, derivative lasso requires the selection of a **single credibility-based parameter**:

the **smoothness**





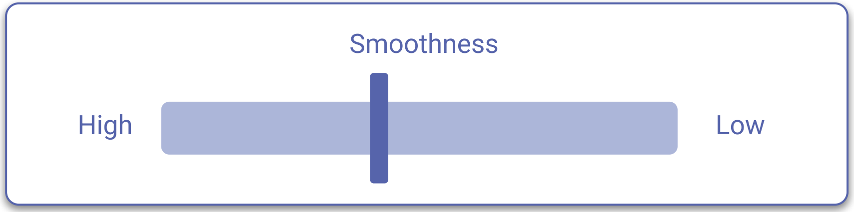
Derivative LASSO

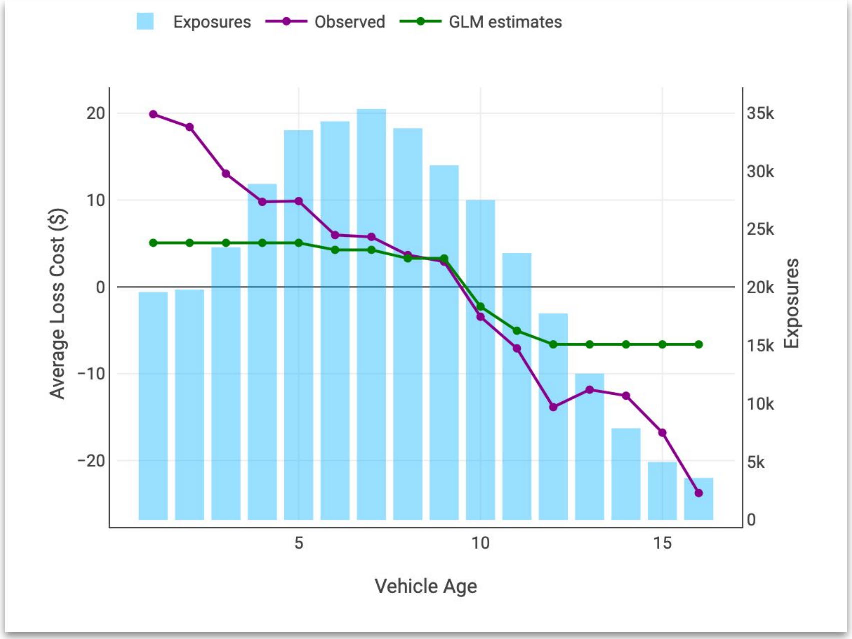
Influence of the smoothness

Derivative lasso **natively incorporates non-linear effects**.

To control the robustness and credibility, derivative lasso requires the selection of a **single credibility-based parameter**:

the **smoothness**





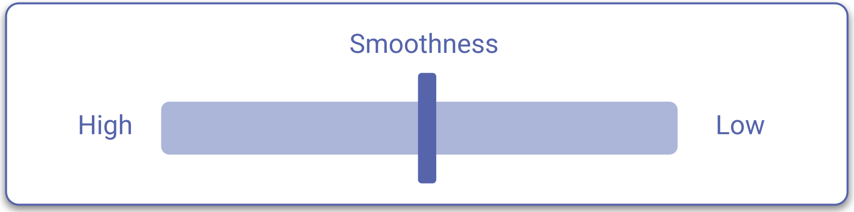
Derivative LASSO

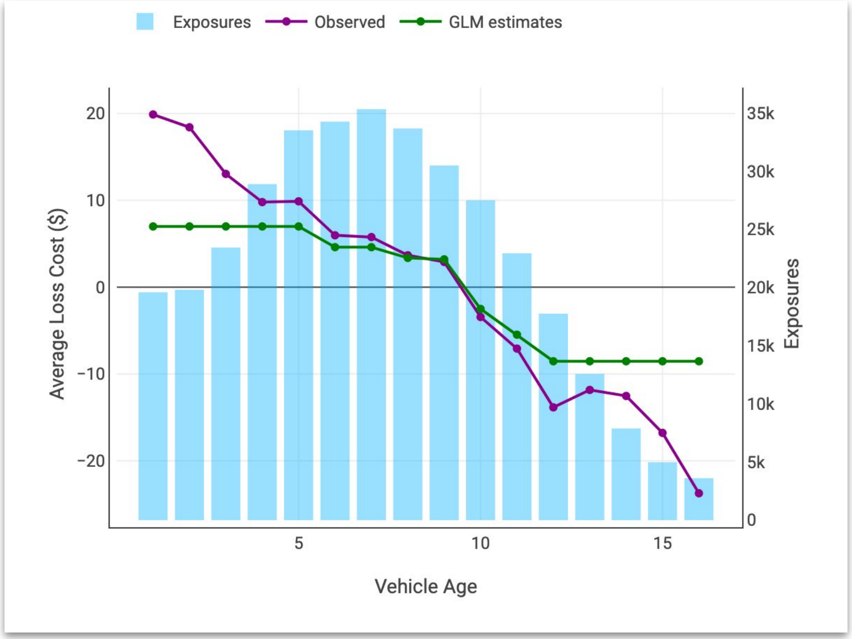
Influence of the smoothness

Derivative lasso **natively incorporates non-linear effects**.

To control the robustness and credibility, derivative lasso requires the selection of a **single credibility-based parameter**:

the **smoothness**





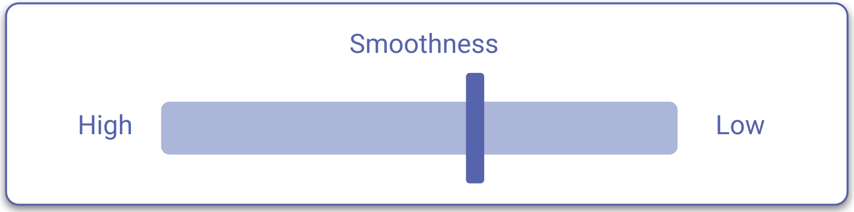
Derivative LASSO

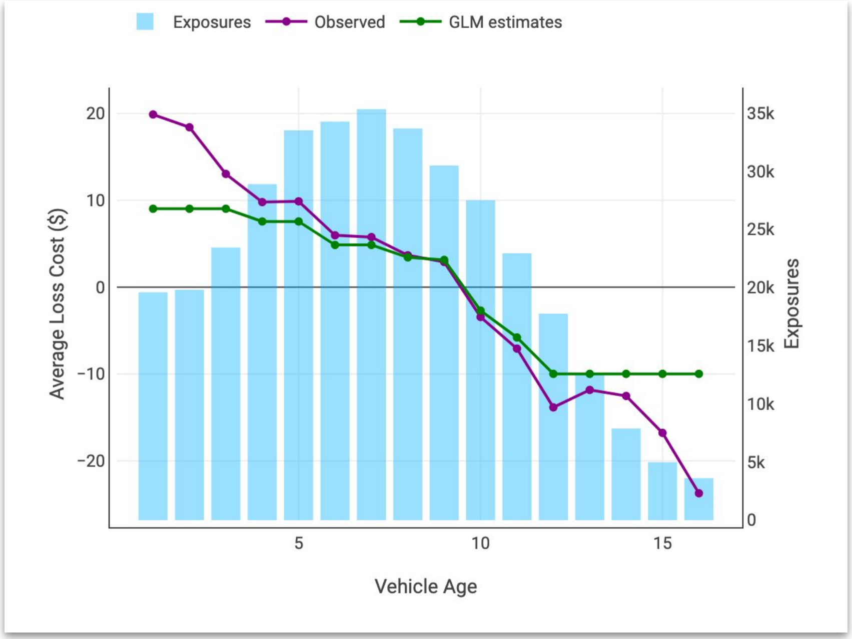
Influence of the smoothness

Derivative lasso **natively incorporates non-linear effects**.

To control the robustness and credibility, derivative lasso requires the selection of a **single credibility-based parameter**:

the **smoothness**





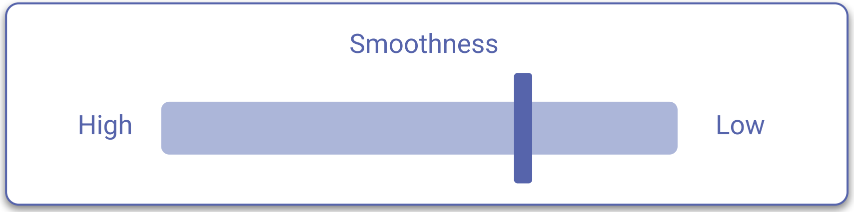
Derivative LASSO

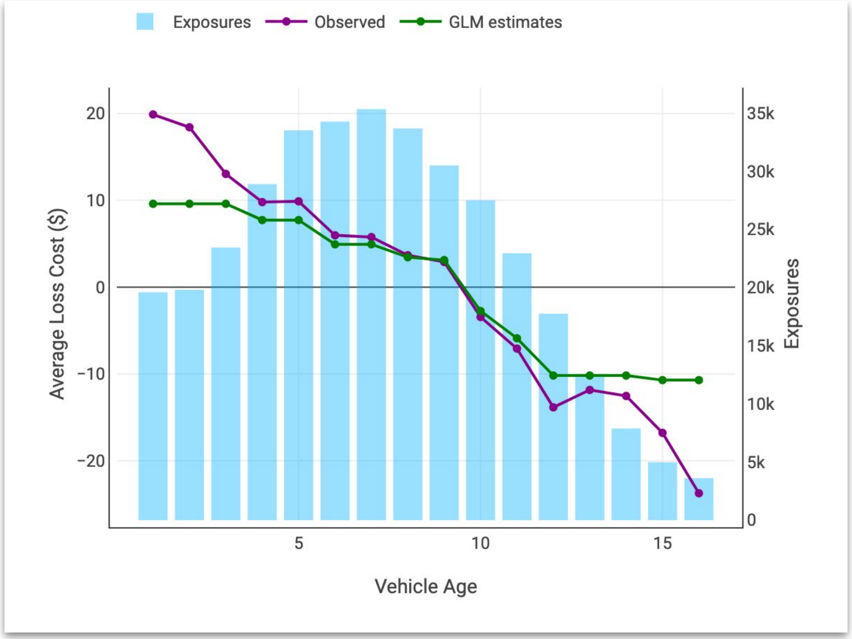
Influence of the smoothness

Derivative lasso **natively incorporates non-linear effects**.

To control the robustness and credibility, derivative lasso requires the selection of a **single credibility-based parameter**:

the **smoothness**





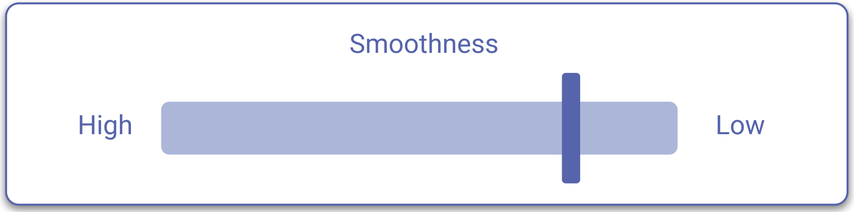
Derivative LASSO

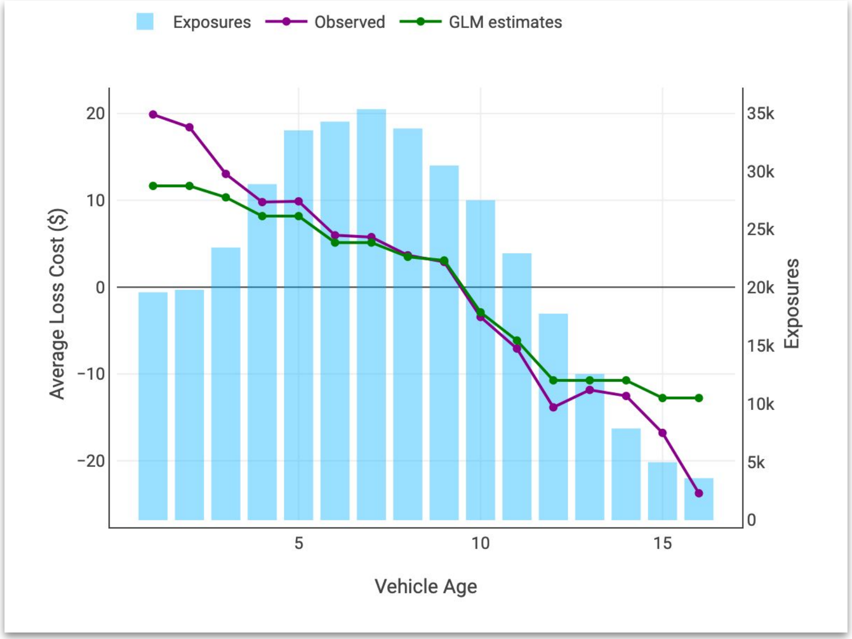
Influence of the smoothness

Derivative lasso **natively incorporates non-linear effects**.

To control the robustness and credibility, derivative lasso requires the selection of a **single credibility-based parameter**:

the **smoothness**





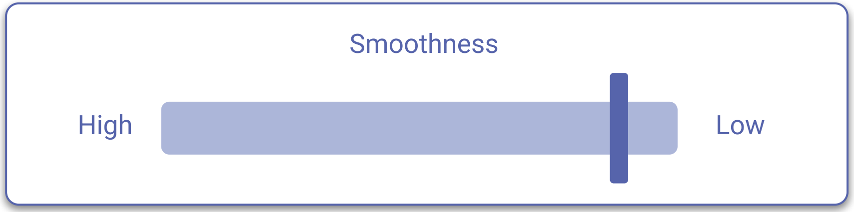
Derivative LASSO

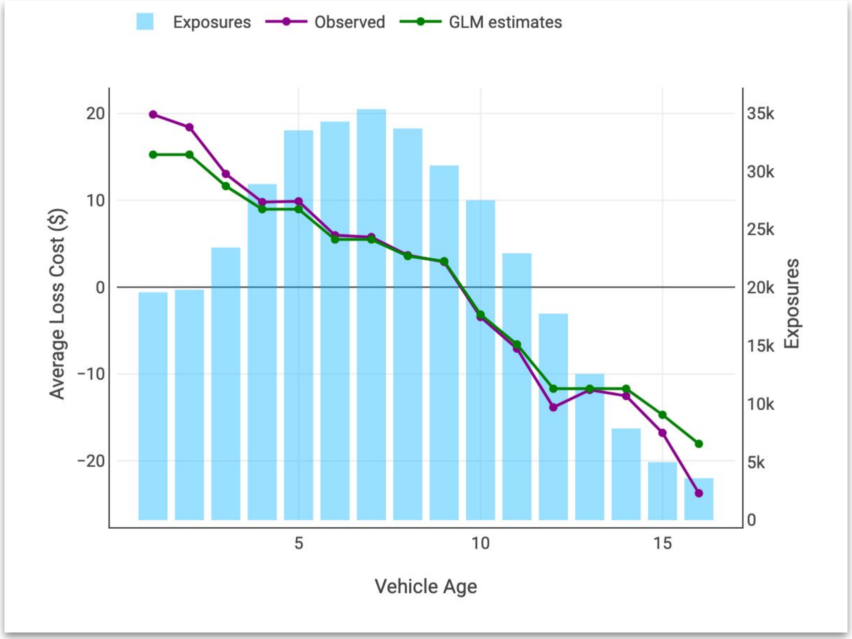
Influence of the smoothness

Derivative lasso **natively incorporates non-linear effects**.

To control the robustness and credibility, derivative lasso requires the selection of a **single credibility-based parameter**:

the **smoothness**





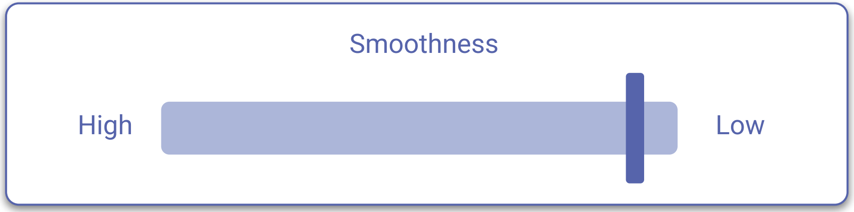
Derivative LASSO

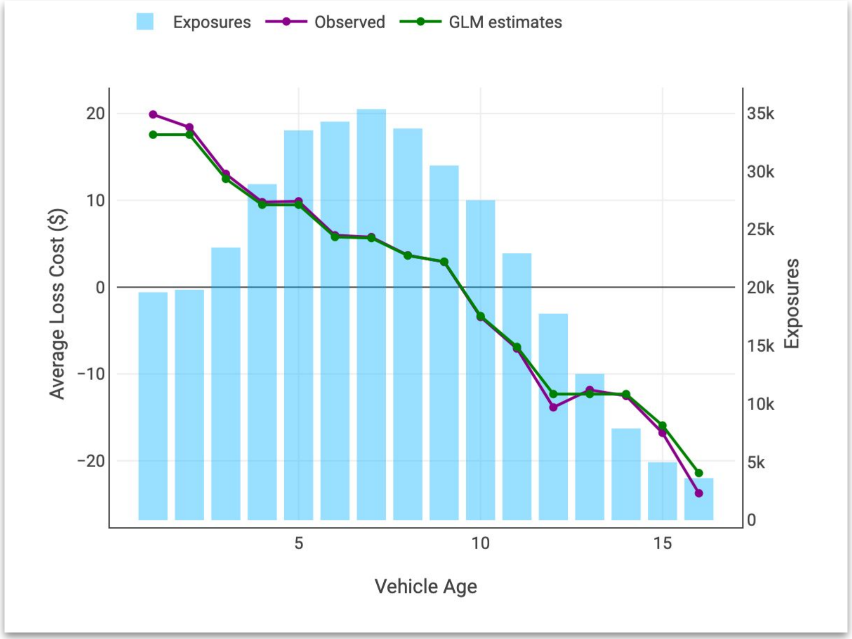
Influence of the smoothness

Derivative lasso **natively incorporates non-linear effects**.

To control the robustness and credibility, derivative lasso requires the selection of a **single credibility-based parameter**:

the **smoothness**



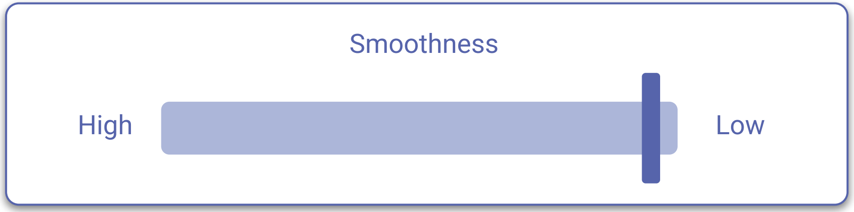


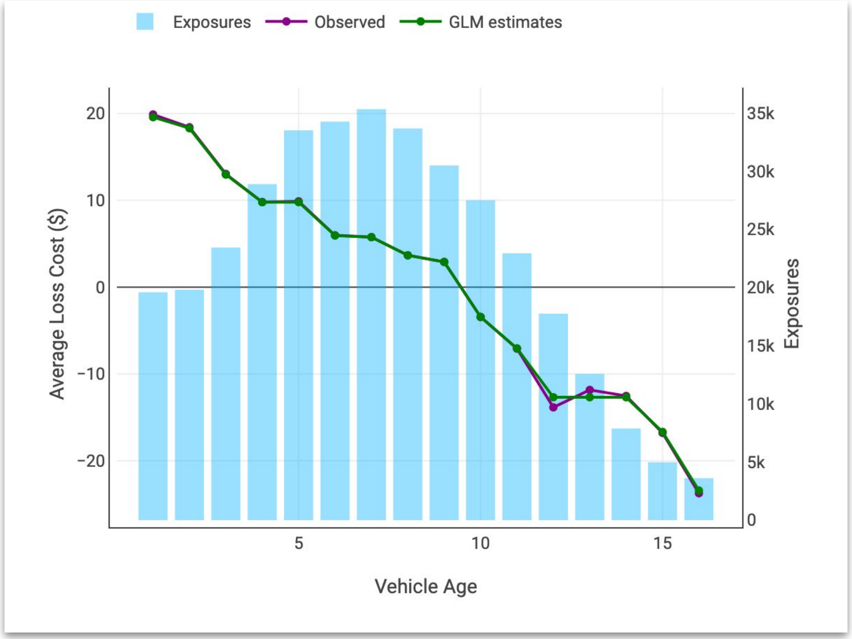
Derivative LASSO
Influence of the smoothness

Derivative lasso **natively incorporates non-linear effects**.

To control the robustness and credibility, derivative lasso requires the selection of a **single credibility-based parameter**:

the **smoothness**



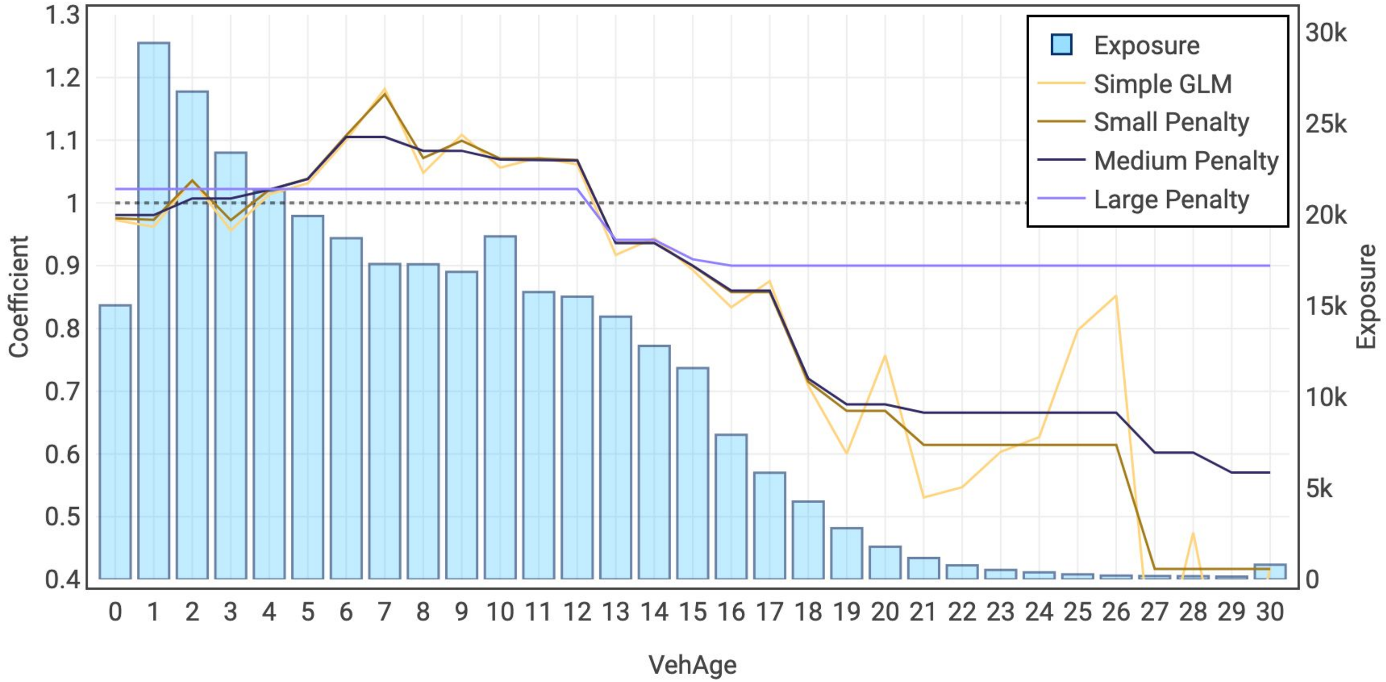


AKUR8

CONFIDENTIAL

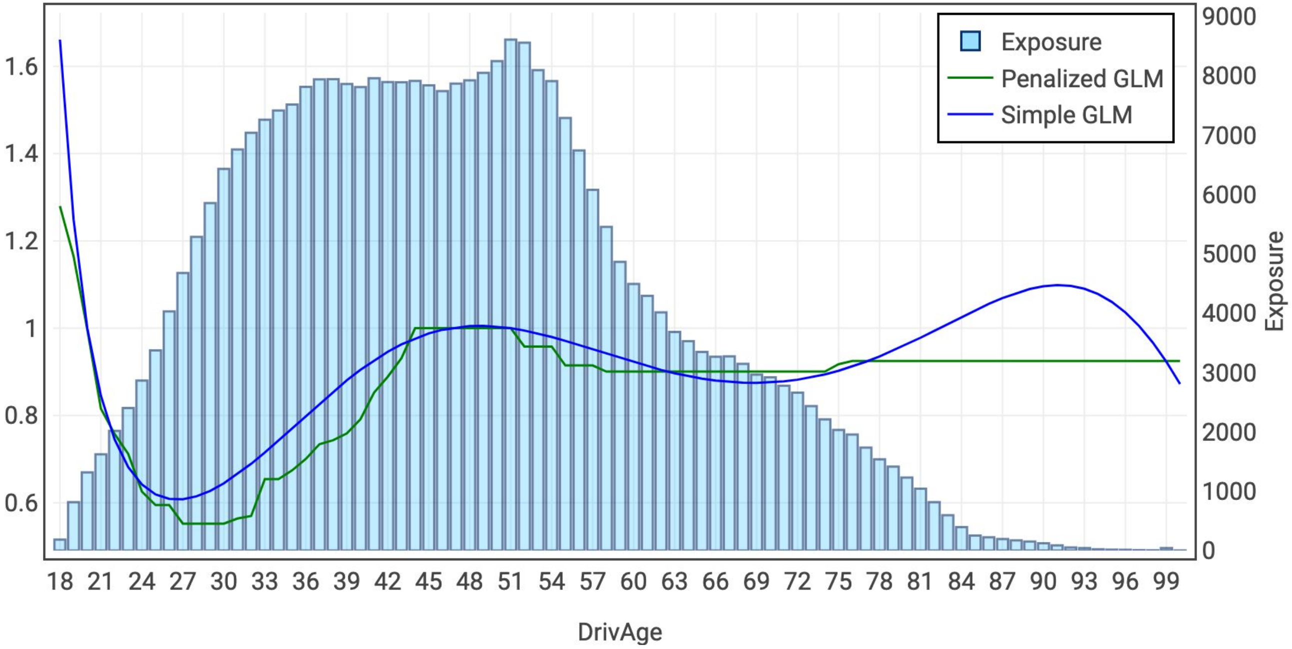
Derivative LASSO

Different values of penalty correspond to different binnings



Derivative LASSO

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



How to select lambda

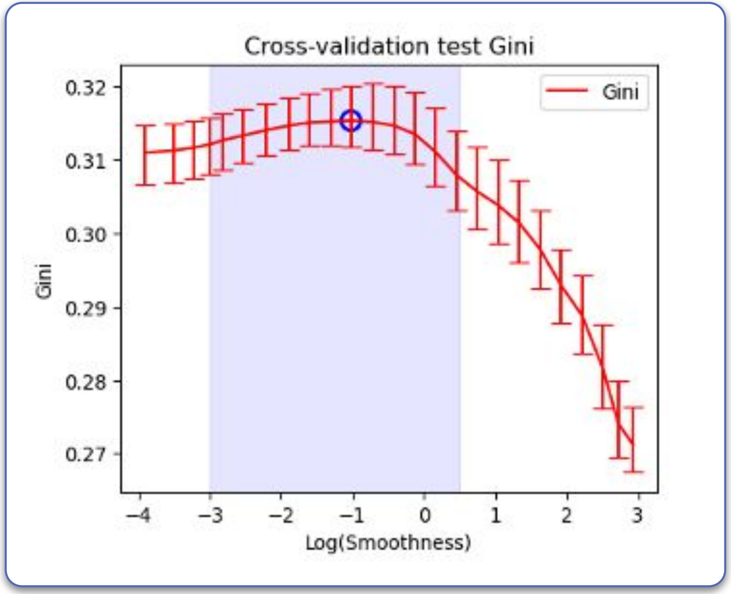
Considerations on Penalized Regression

Hyperparameter tuning - I can find the optimal λ through a cross validation

Performance:
vertical y axis

The higher - the better

Performance is an approximation of the performance on unseen data - measured via **cross validation.**

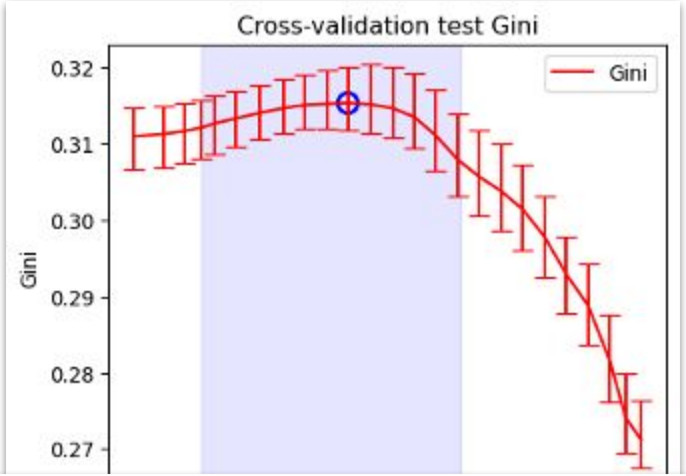


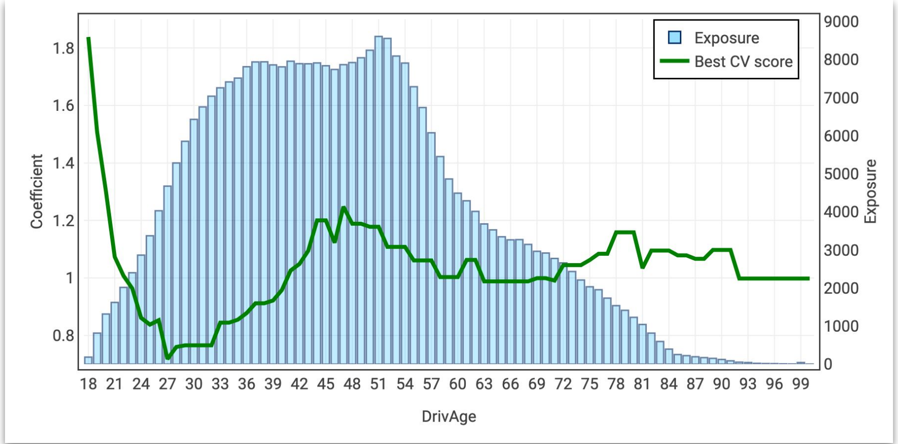
Smoothness of models ranging from **low (equal to GLM) to high**.

Performance is not all you need

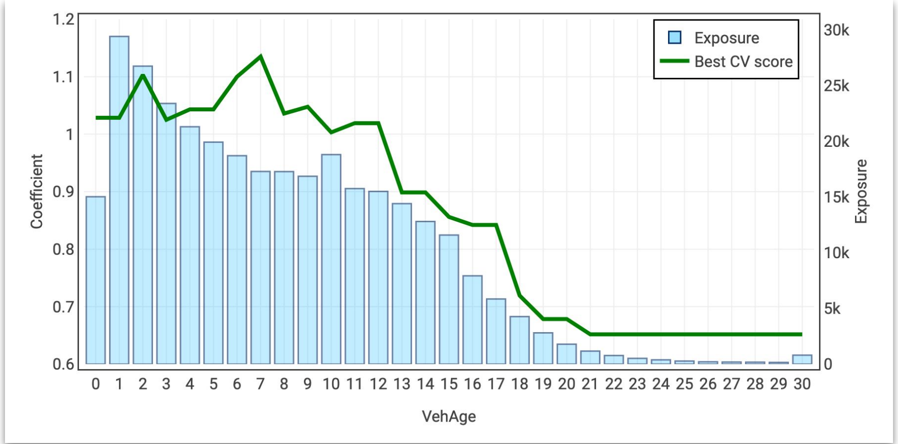
The best performing model may detect undesired signal

The **best performing** model has **too many reversals.**





(a)

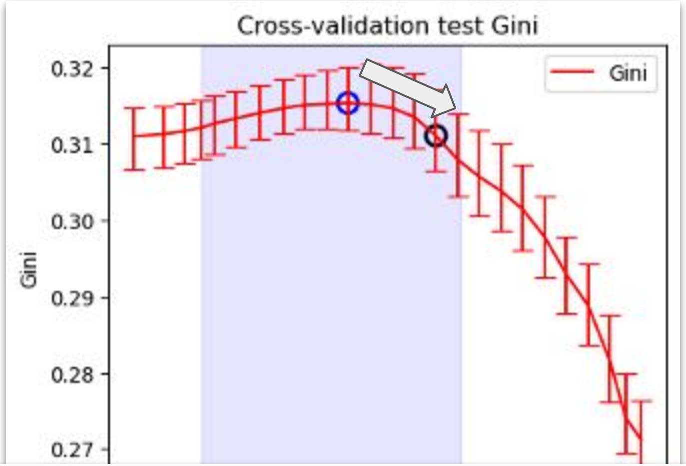


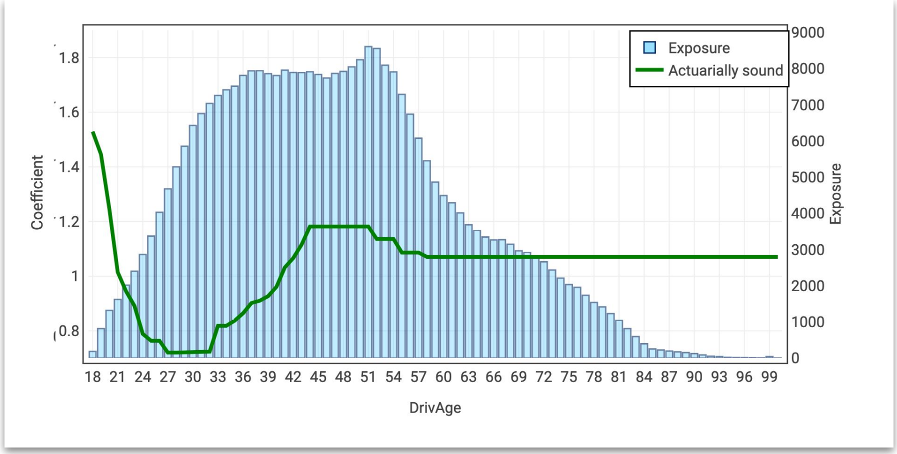
(b)

Second phase: Actuarial judgment

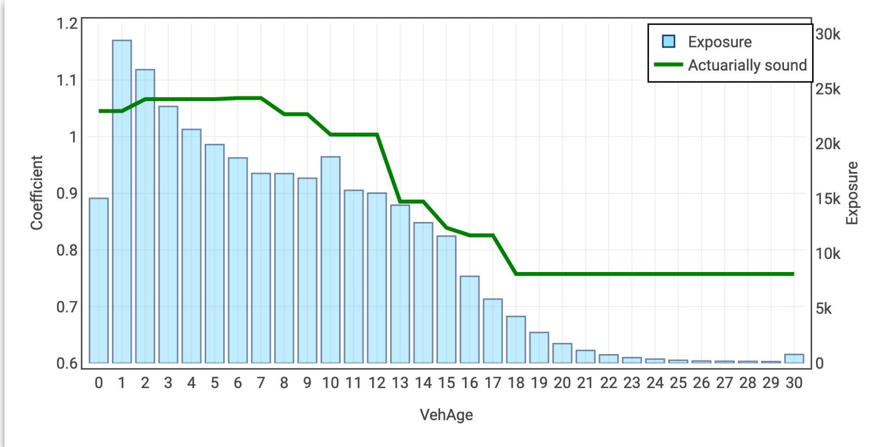
Use actuarial judgment to select the smoothness

Ability to **control smoothness** parameter leads to **sound and robust estimates**.





(a)



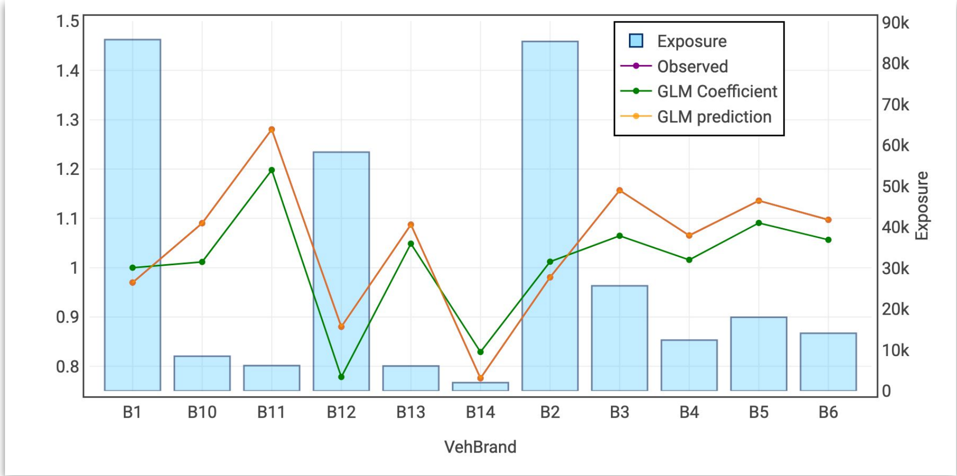
(b)

Conclusions

Considerations on Penalized Regression

Penalized Regression improve the performance of GLMs keeping the transparency

- **Performance**
  - Suitable for **high dimensionality** use-cases (a lot of variables) **preventing overfitting**
  - **Automatic feature selection**
- **Interpretability**
  - Same structure of GLM: **high interpretability**
  - Generalization of the **credibility** framework
- **Tuning**
  - The **penalization** (smoothness) **can be tuned** by the modeler
  - Possibility to set **monotonicity constraints** for ordinal variables
  - Full **control** on **interactions**
  - The modeler can **edit manually** the coefficients in a GLM fashion



Q&A time!

Leonardo.Stincone@akur8.com
linkedin.com/in/leonardo-stincone/

Mattia.Casotto@akur8.com
linkedin.com/in/mattia-casotto/

THANKS

28 rue de Londres, 75009 Paris
FRANCE

AKUR8