An Introduction to Derivative Lasso and Lasso Credibility

2023

Agenda

NAIC Book Club: October 24, 2023

Introduction to Derivative Lasso
Smoothness and Credibility
Selecting the Smoothness
Items for Review: Derivative Lasso
Lasso Credibility
Items for Review: Lasso Credibility

Derivative Lasso

Akur8's Modeling Methodology: Derivative Lasso

Demystification

Derivative Lasso

Penalty on Change

One tuning parameter

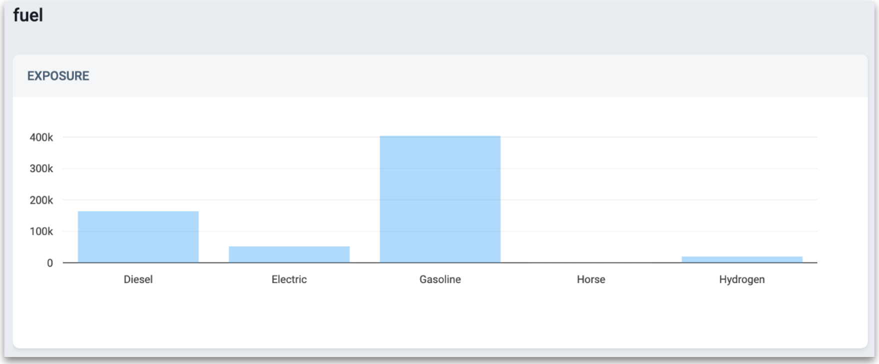
No user-input variable transformations

Automatically removes noncredible coefficients

Derivative Lasso: Variable Inputs are Ordinal or Categorical

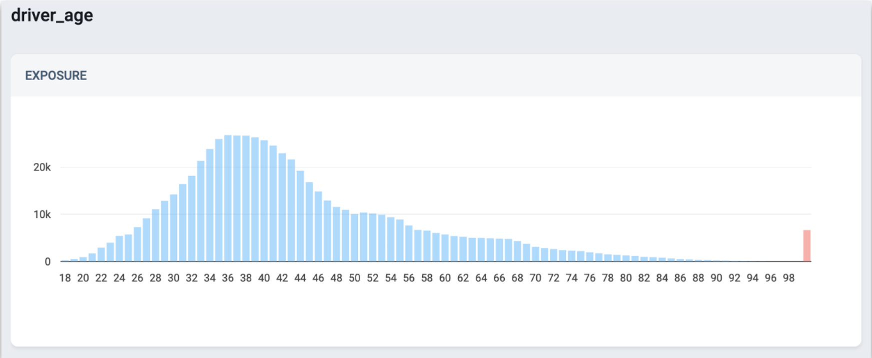
Continuous variables are treated as ordinal

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.

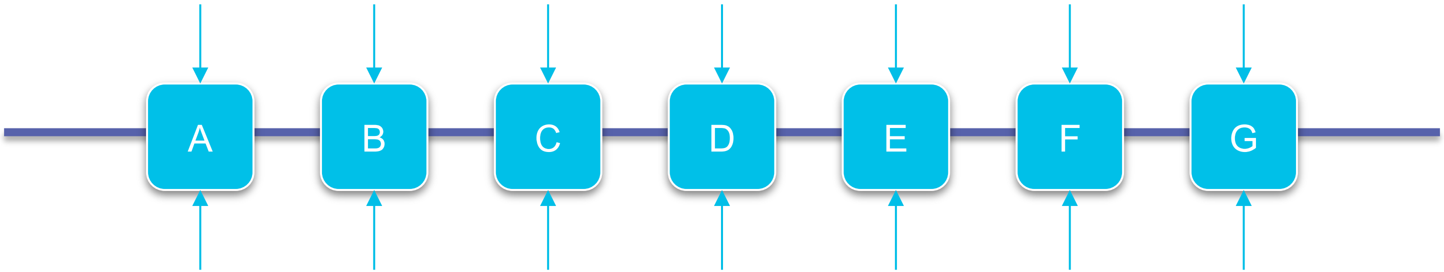
Could also be A, B, C, D, E….



One Tuning Parameter: The Smoothness Penalty

Not mgcv smoothness

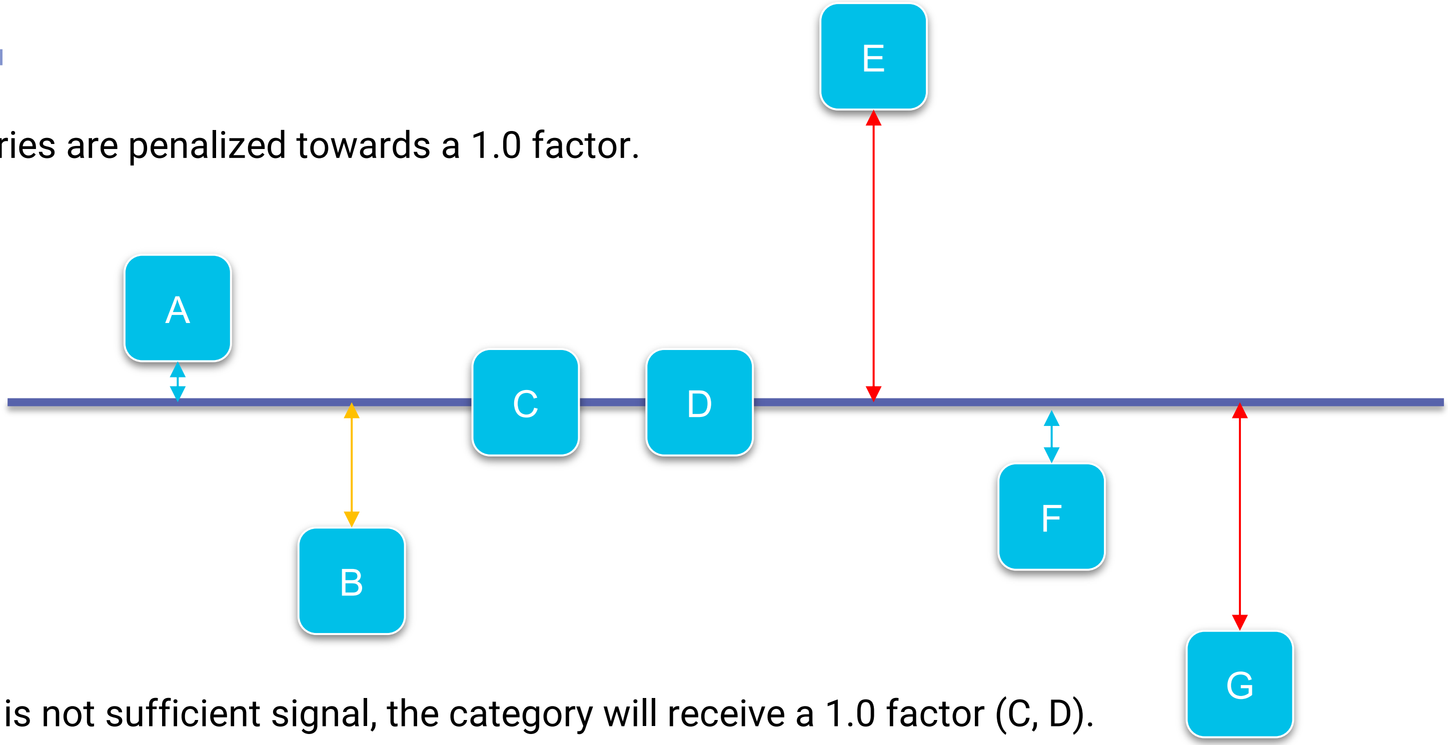
Categories are penalized towards a 1.0 factor.



One Tuning Parameter: Smoothness

Not mgcv smoothness

Categories are penalized towards a 1.0 factor.



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

One Tuning Parameter: Smoothness

Not mgcv smoothness

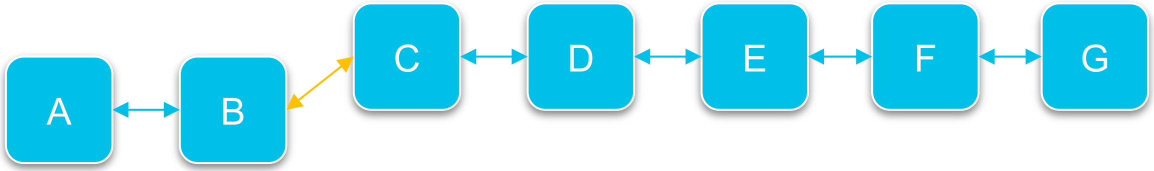
The difference (derivative) between adjacent levels is penalized for ordinal variables.

Levels without sufficiently different signal are grouped together.

One Tuning Parameter: Smoothness

Not mgcv smoothness

With a large smoothness penalty, only the most credible differences will be identified.

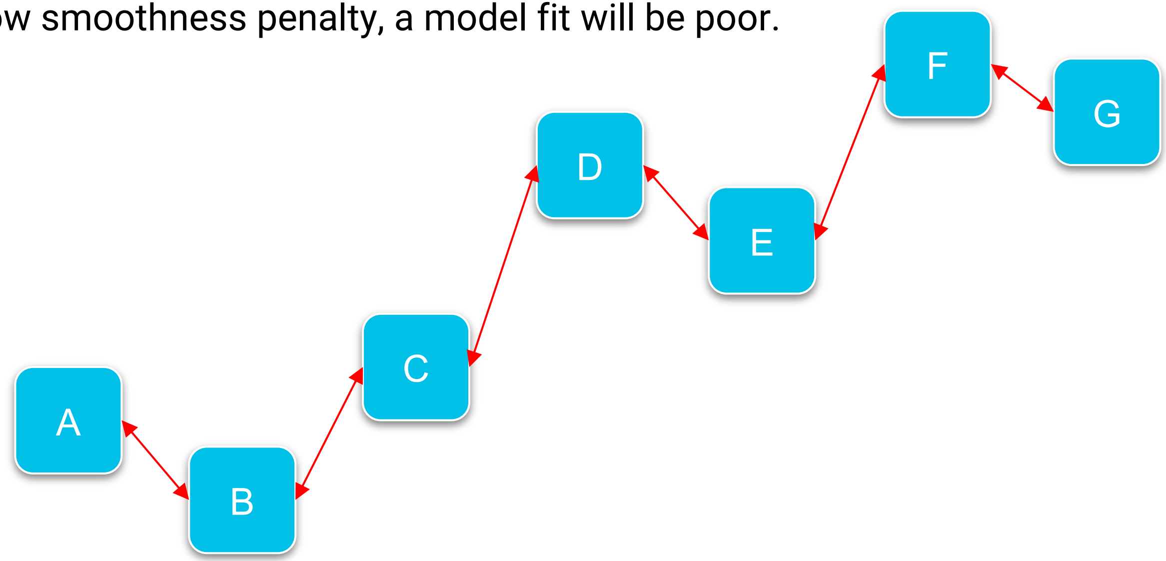


This model is **too smooth**.

One Tuning Parameter: Smoothness

Not mgcv smoothness

With a far too low smoothness penalty, a model fit will be poor.

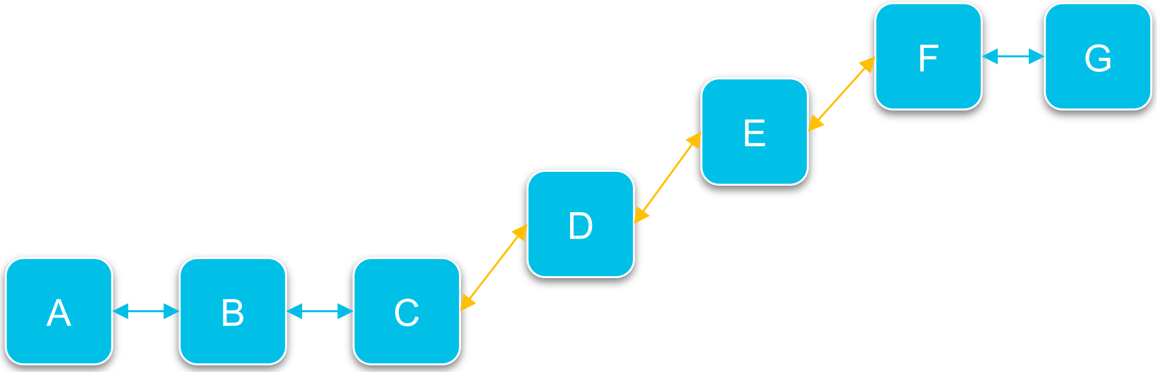


This model is **not smooth enough**.

One Tuning Parameter: Smoothness

Not mgcv smoothness

With an appropriate derivative lasso penalty, the differences will be penalized to create an intuitive fit.



This model has an appropriate **amount of smoothness**.

Why Derivative Lasso a GAM?

A different kind of GAM

The basis function of derivative lasso divides ordinal variables into individual levels. A penalty is applied between each of these many levels.

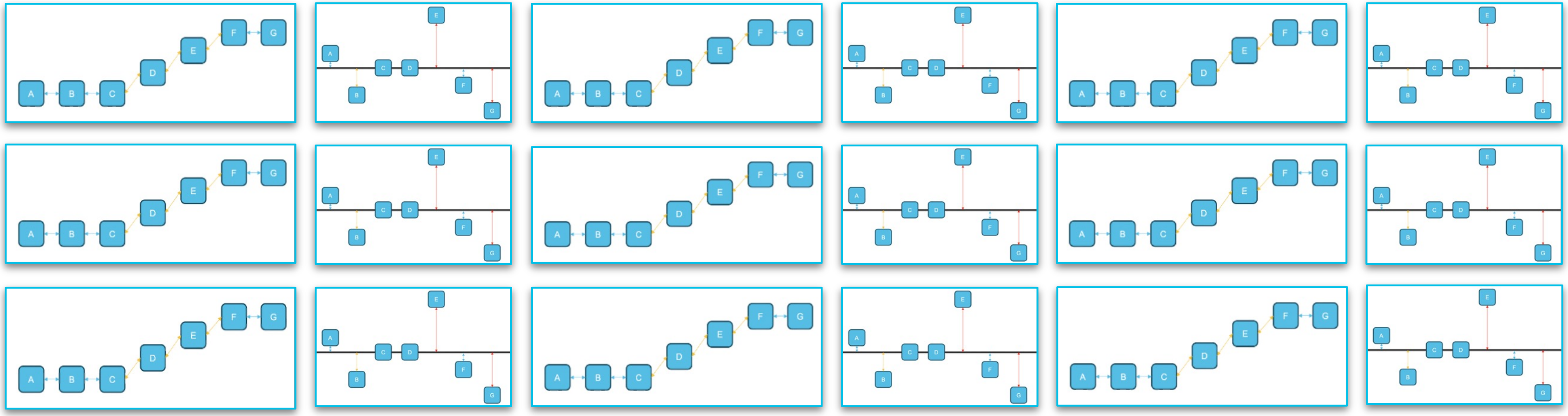
10,000 ⟷ 10,001 ⟷ 10,002 ⟷ … ⟷ 100,000 ⟷ 100,001 ⟷ 100,002 ⟷ … ⟷ 980,002..

Applications of derivative lasso may group levels to reduce computation time.

Rather than splines or hinges, **a derivative lasso basis function is purely stepwise** and highly intuitive.

One Tuning Parameter: Smoothness

Not mgcv smoothness



The smoothness parameter determines all factor levels and curves simultaneously.

Tuning is not allowed on a variable-by-variable basis, as the smoothness is a credibility standard applied simultaneously to all categorical levels and differences between ordinal levels in the model.

Smoothness is a Credibility Procedure

Smoothness is a Lasso Penalty – Linked to Credibility

Literature exists to link (and sometimes equate) penalized regression and credibility

Current and Future Publications:

- Derivative Lasso: Credibility-Based Signal Fitting for GLMs
- Penalized Regression and Lasso Credibility *(Future CAS Monograph in Review)*
- Credibility and Penalized Regression
- A Discussion on Credibility and Penalised Regression, with Implications for Actuarial Work

2022-2023 Presentations at CAS, iCAS, NAIC, and SOA Events:

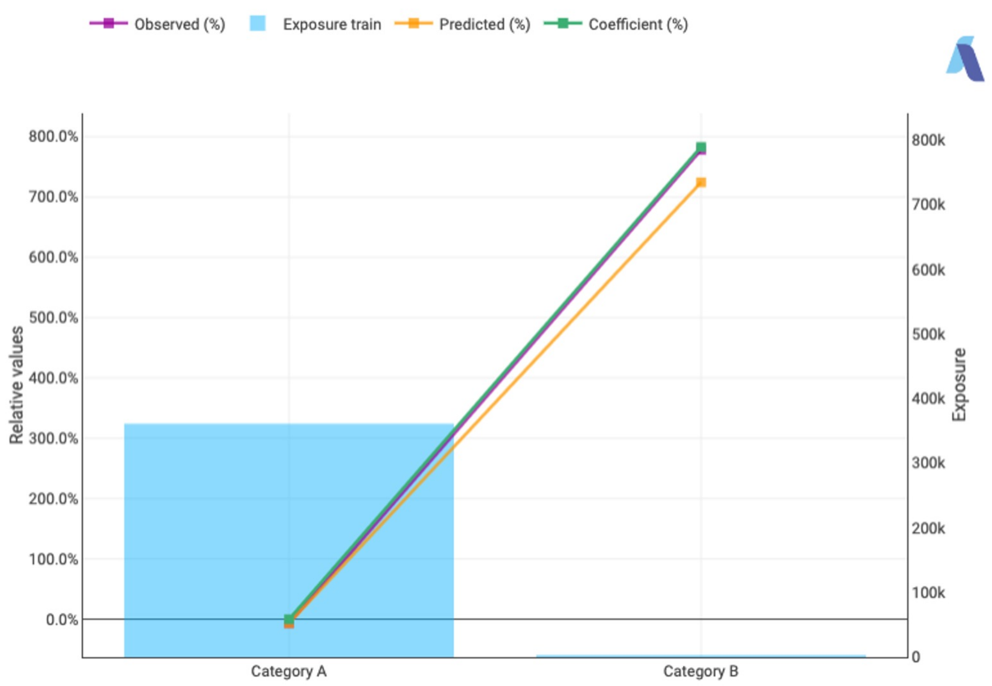
- Penalized Regression - Between Credibility and GBMs
- Applying Credibility in Penalized Regression
- Penalized Regression and Lasso Credibility
- Derivative Lasso (November Annual Meeting 2023)

Papers contain mathematical proofs for those interested, presentations require various levels of technical knowledge.

Penalizing a One-Variable Model

Penalization is credibility

Fully Credible GLM:
Y = Intercept + Bx

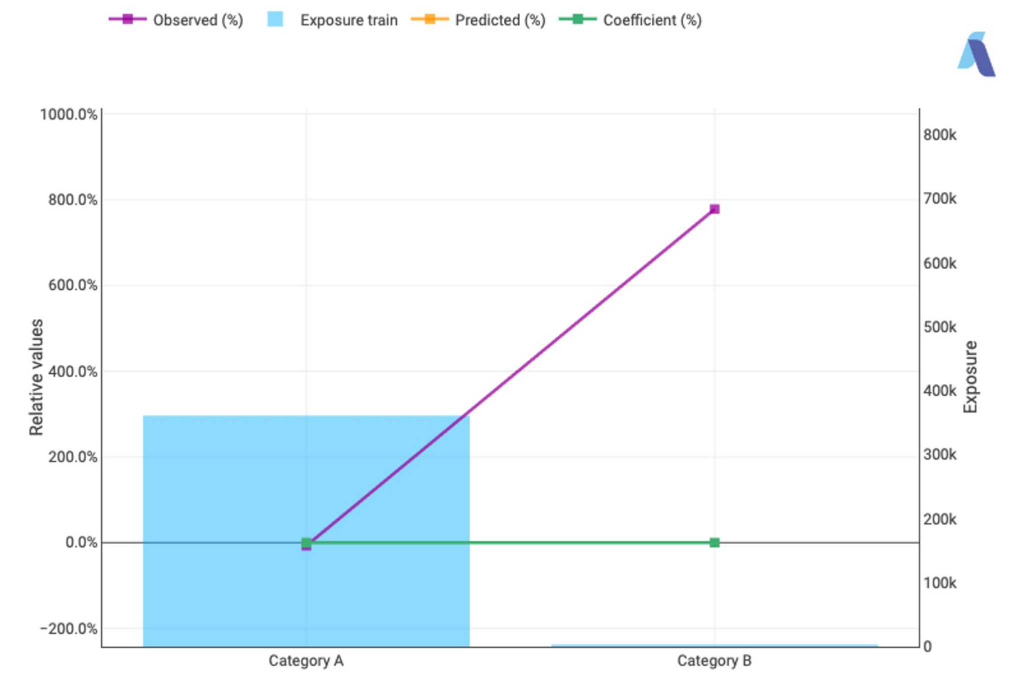


Penalizing a One-Variable Model

Penalization is credibility

Fully Credible GLM:
Y = Intercept + Bx

Fully Penalized GLM:
Y = Intercept + 0x



Penalizing a One-Variable Model

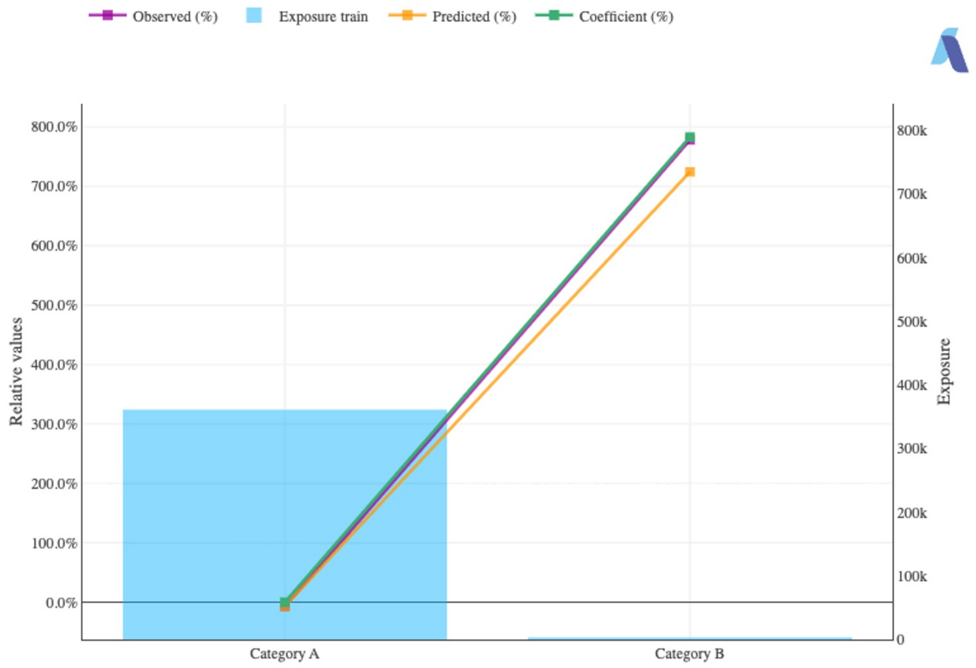
Penalization is credibility

Fully Credible GLM:
Y = Intercept + Bx

Fully Penalized GLM:
Y = Intercept + 0x

Partially Penalized GLM:
Y = Intercept + B*x
B > B* > 0

$$\beta_i = \beta_{GLM} \times (Z) + 0 \times (1 - Z)$$



Statistical Link: Credibility and Penalized Regression

Cross validation and penalization combine to create a robust multivariate credibility procedure

Ridge penalized regression's penalty is **equivalent to the Buhlmann K** in a univariate special case.*

$$\lambda = K = \frac{\tau^2}{\sigma^2}$$

The lambda penalty parameter's **credibility behavior generalizes** to multivariate modeling.

$$\arg\min_{\beta} -\text{LogLikelihood}\left(y_i, X, \beta\right) + \lambda\, \text{Penalty}(\beta)$$

Penalized

Regression (GLM)

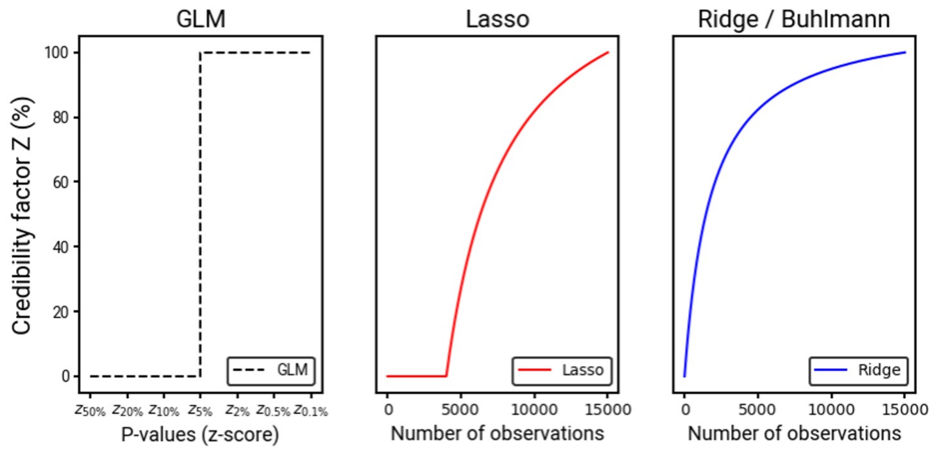
Robust methodology to estimate an appropriate lambda penalty parameter already exists in data science literature and is well accepted: **cross validation**.

Variable Removal: Lasso has a Credibility Threshold

Cross validation and penalization combine to create a robust process for removal

Lasso penalization favors **sparsity**, or model simplicity.

Sparsity is usually preferred for actuarial analysis.



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

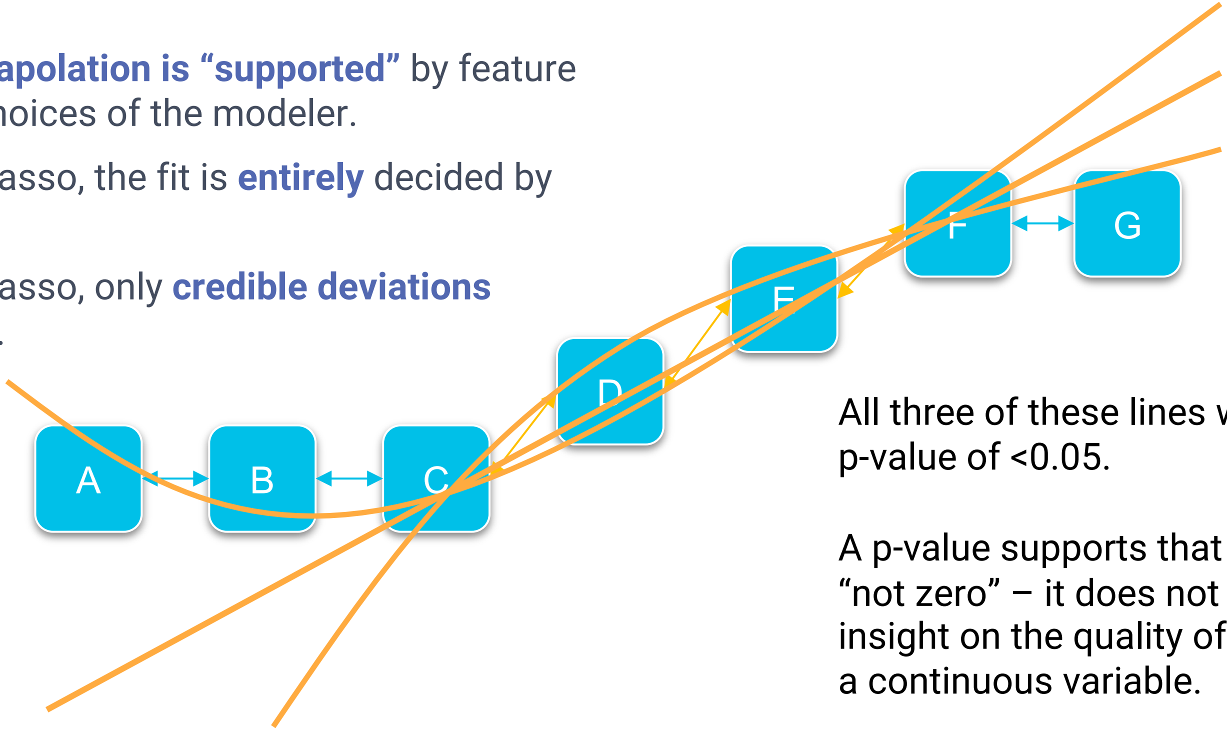
An Entirely Credibility-Based Fitting Procedure

Credibility replaces feature engineering

In a GLM, **extrapolation is “supported”** by feature engineering choices of the modeler.

In Derivative Lasso, the fit is **entirely** decided by the data.

In Derivative Lasso, only **credible deviations** are supported.

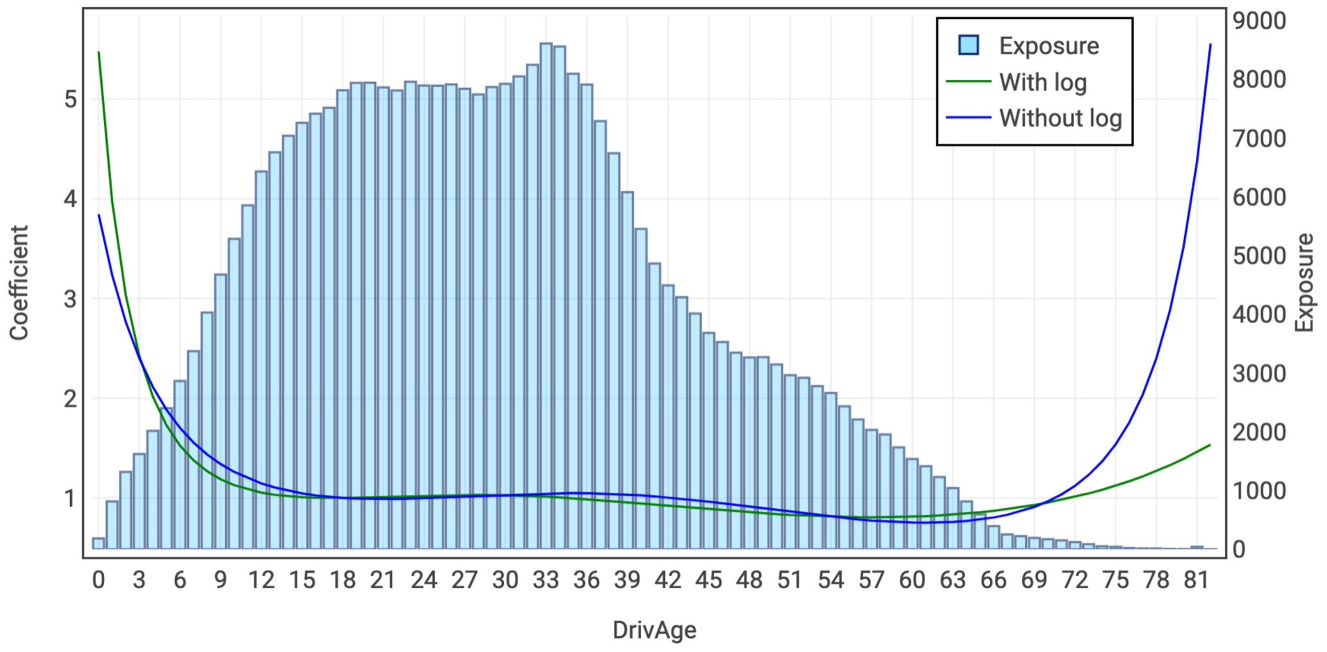


All three of these lines would have a p-value of <0.05.

A p-value supports that the slope is “not zero” – it does not provide insight on the quality of the slope of a continuous variable.

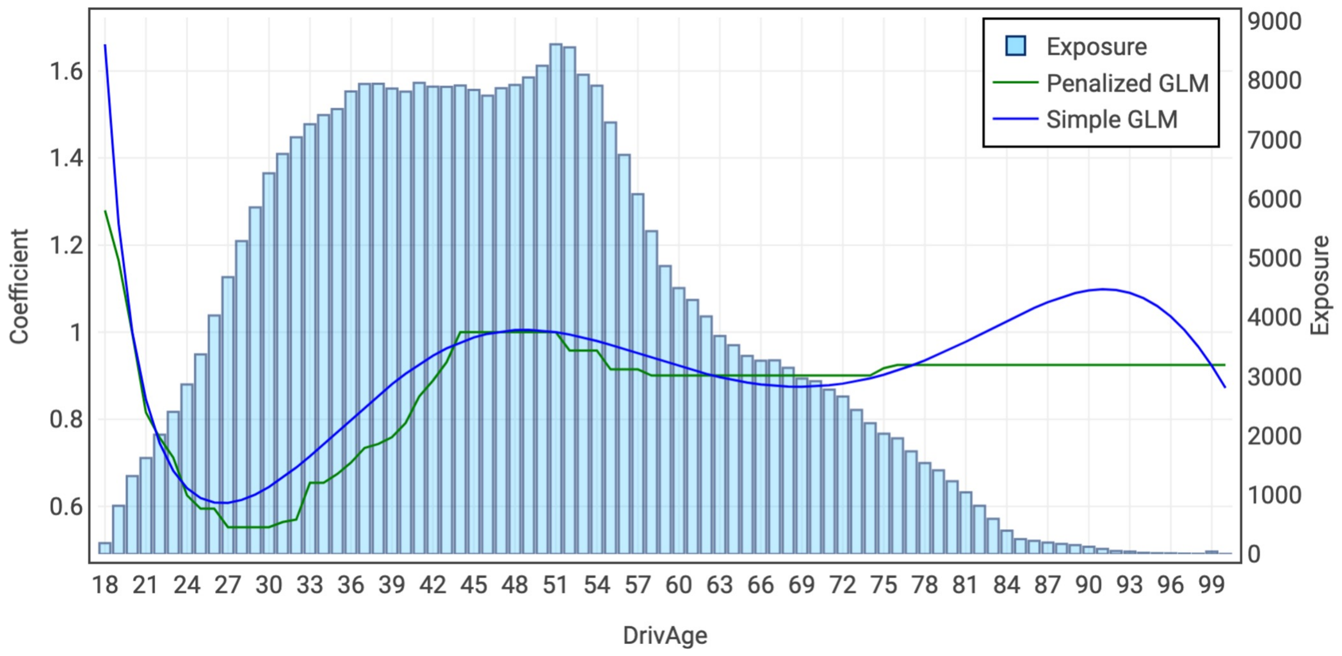
Extreme Example of GLM's "Support" for Extrapolation

All GLM coefficients in this example have p-values below 0.05



Derivative Lasso Automatically Captures the Signal from the Data

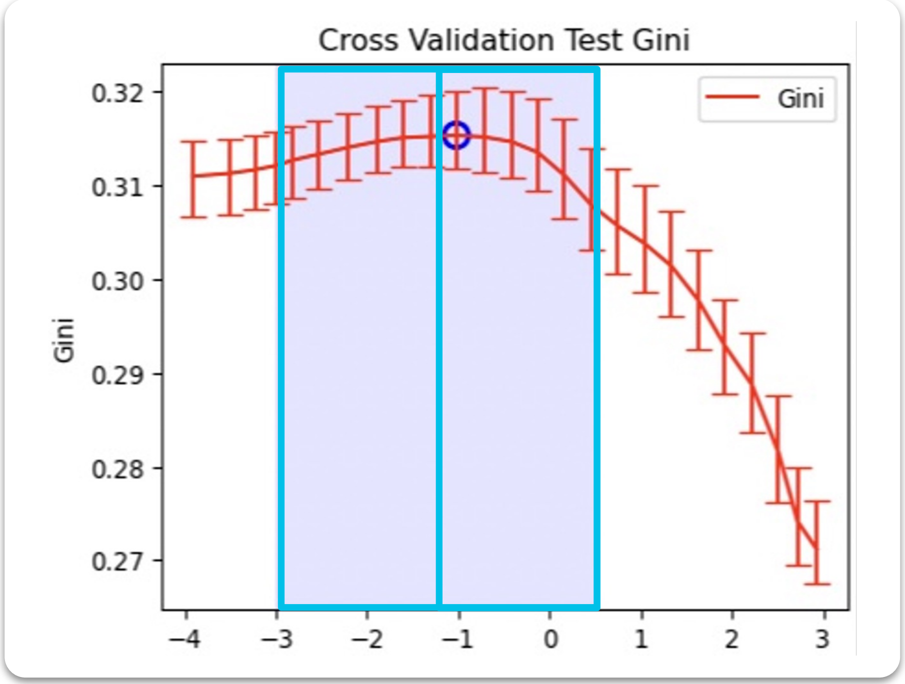
All simple GLM coefficients have p-values below 0.05. Penalized GLM is derivative lasso



Tuning the Smoothness

A Reasonable Range of Penalty Terms

Incorporating judgment through the choice of smoothness



Cross validation provides a range of statistically reasonable penalty terms. Many of these penalty terms will produce models with very similar statistical accuracy.

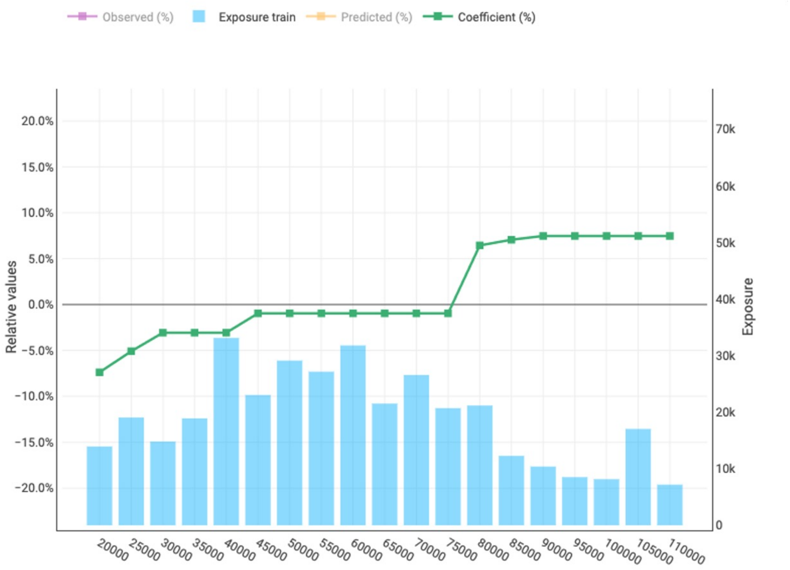
Using cross validation to select a penalty term is thoroughly supported in the literature.

As an **actuarial best practice**, we recommend only the less reactive side of this reasonable range. Exceptions may exist.

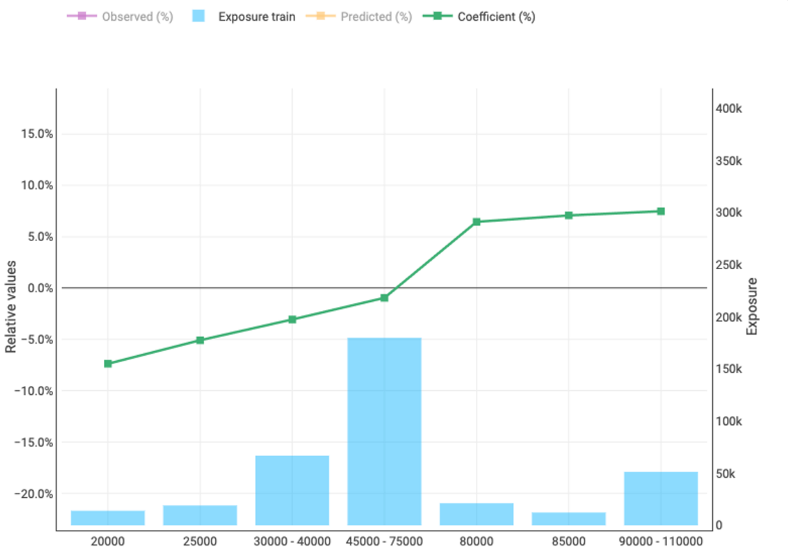
Example Univariate Plots – Vehicle Value

Two equivalent graphs of appropriate smoothness

Derivative Lasso Plot – Raw Levels



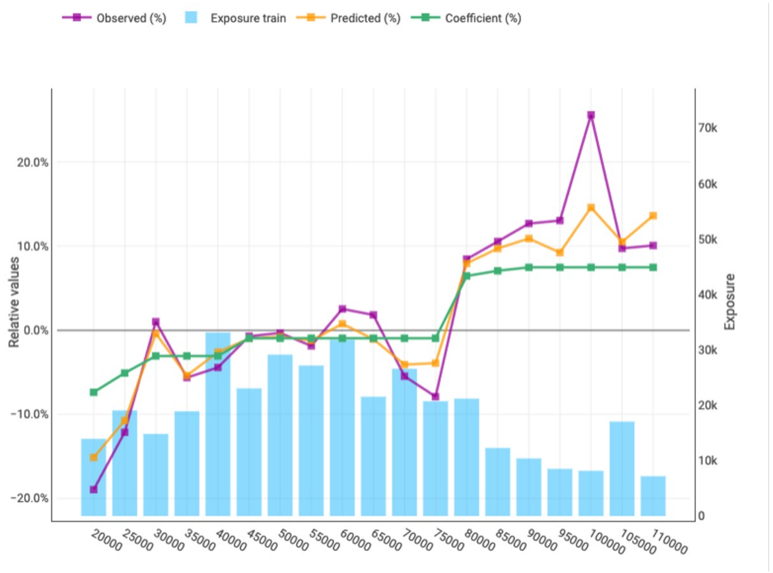
Derivative Lasso Plot – Binned Levels



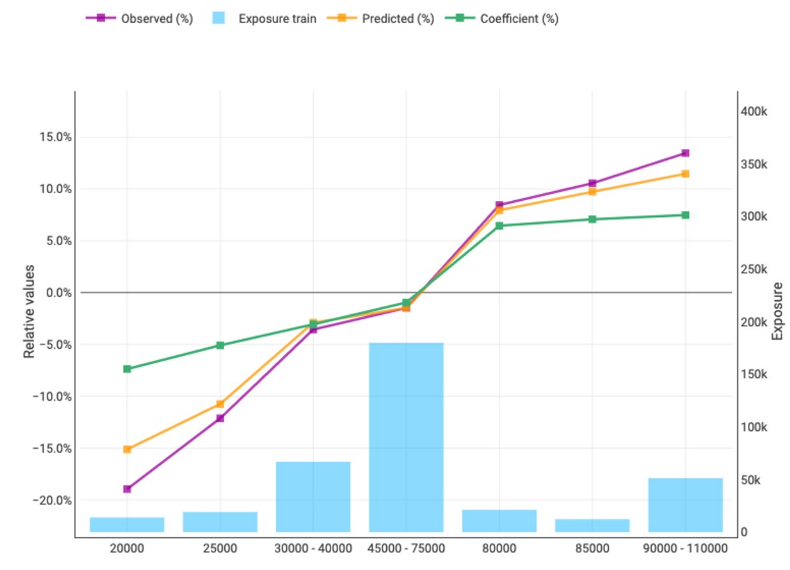
Example Univariate Plots – Vehicle Value

Two equivalent graphs of appropriate smoothness

Derivative Lasso Plot – Raw Levels



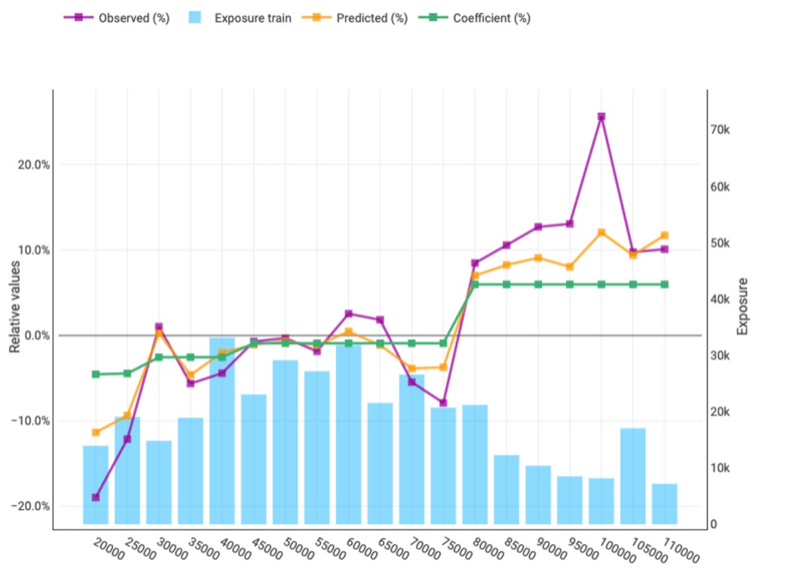
Derivative Lasso Plot – Binned Levels



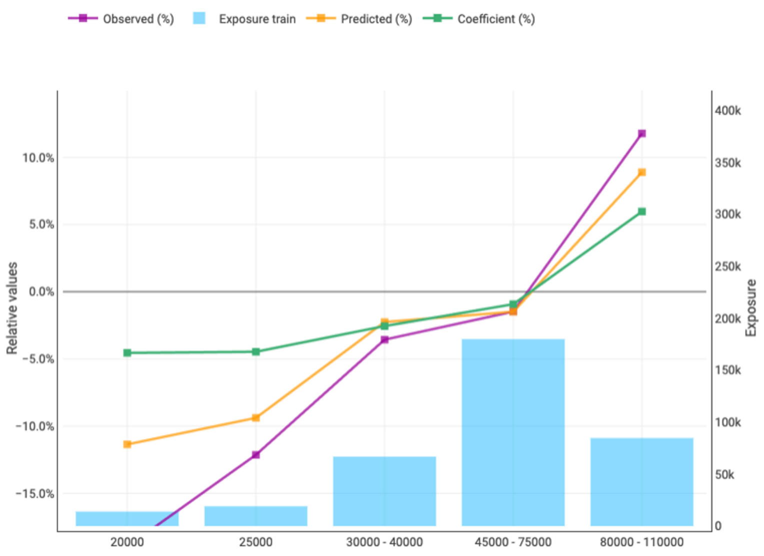
Example Univariate Plots – Vehicle Value

Two equivalent graphs of high smoothness

Derivative Lasso Plot – Raw Levels



Derivative Lasso Plot – Binned Levels

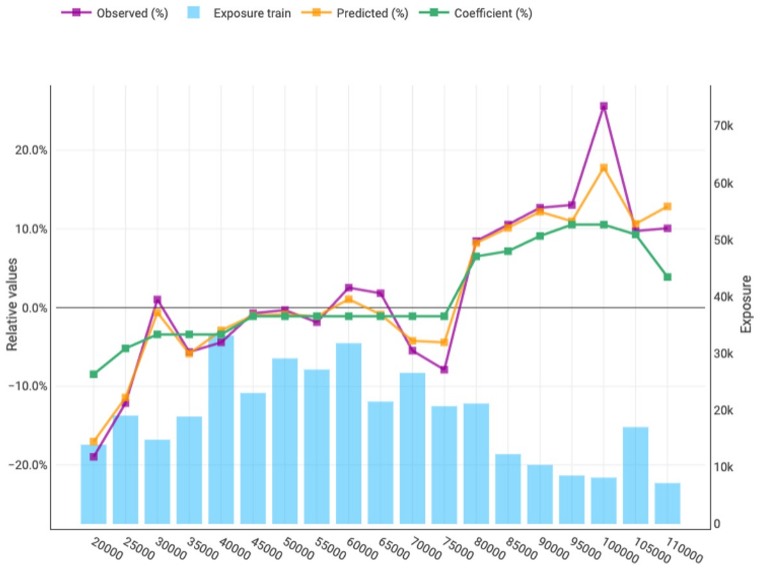


AKUR8

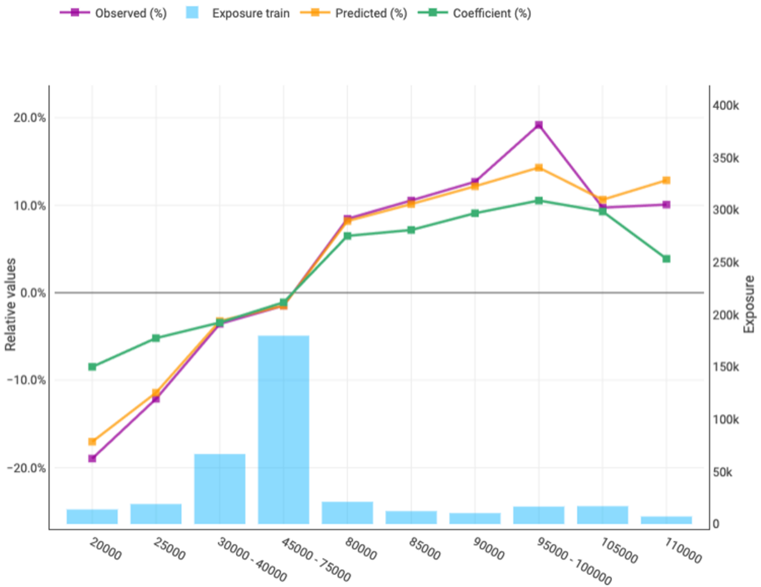
Example Univariate Plots – Vehicle Value

Two equivalent graphs of low smoothness

Derivative Lasso Plot – Raw Levels

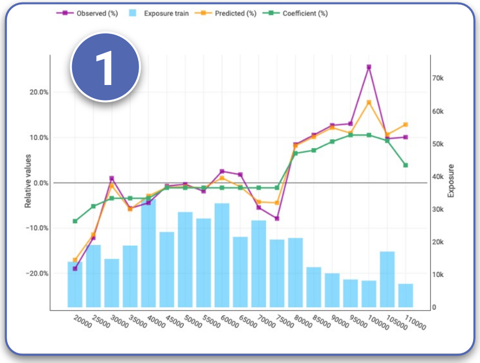


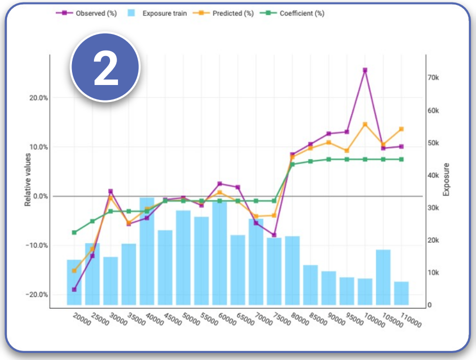
Derivative Lasso Plot – Binned Levels

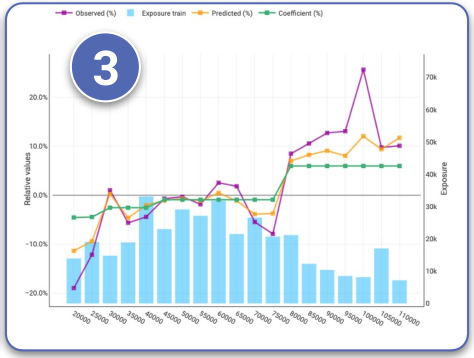


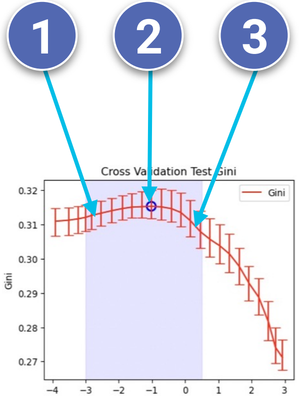
Visualizing Different Smoothness Selections

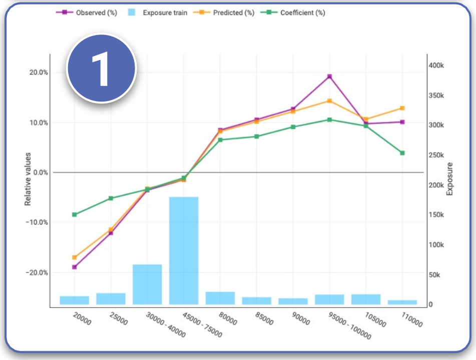
Top and bottom graphs are identical – binning is automated through the smoothness parameter





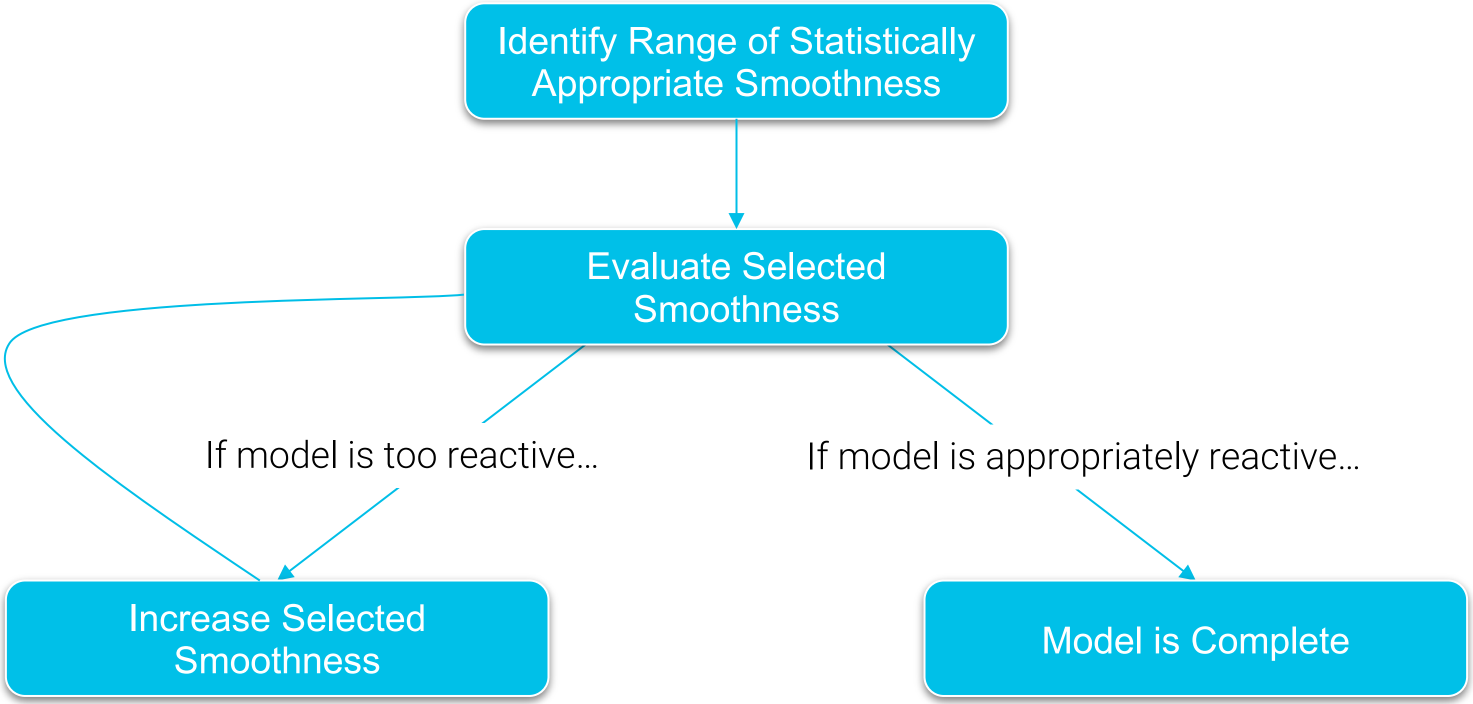






Standard Steps for Building a Derivative Lasso Model

Best case scenario



Common Modeler Questions

Adopting a purely credibility-driven approach

Q: I want this variable in my model, but it is excluded by the penalty. How do I force it into my model?

A: You can't. If the variable's signal doesn't overcome the penalty, the variable is not credible enough and is excluded.

Q: I want to force a variable to strictly monotonically increase.

A: You can, but in most cases, you shouldn't have to. Undesirable reversals are usually caused by data outliers. Correctly adjusting the data for outliers usually removes these unintuitive reversals.

Q: I want a custom binning for my variable.

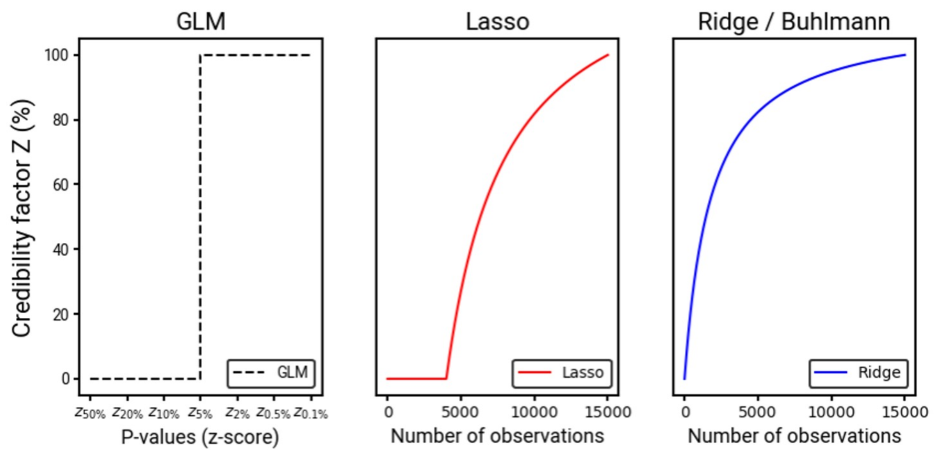
A: Custom binning should be applied only for implementation restrictions or other restrictions, not for a desired model fit.

Items for Review in Derivative Lasso

Items to Assist Review

Typical derivative lasso review items

- Validate Lift Charts.
- Validate Gini Charts.
- Univariate Charts.
- Current, Indicated, Selected by Variable.
- Output coefficients are identical to a GLM.
- **List of post-modeling adjustments**



No universally applicable P-value substitutes are available. Instead, we rely on the selected threshold of credibility for lasso penalized regression.

Watch Items

Derivative lasso models

Watch items are adjustments that aim to “get around” the smoothness parameter.

Because the smoothness replaces judgmental feature engineering and noncredible variables are automatically removed, it is difficult to build a model with hidden technical deficiencies.

Unnecessary pre-grouping of variables to force desired behavior instead of real behavior.

- OK: Grouping to match current rating tables or implementation restrictions.
- Watch: Grouping excessively to allow for a less-smooth model.

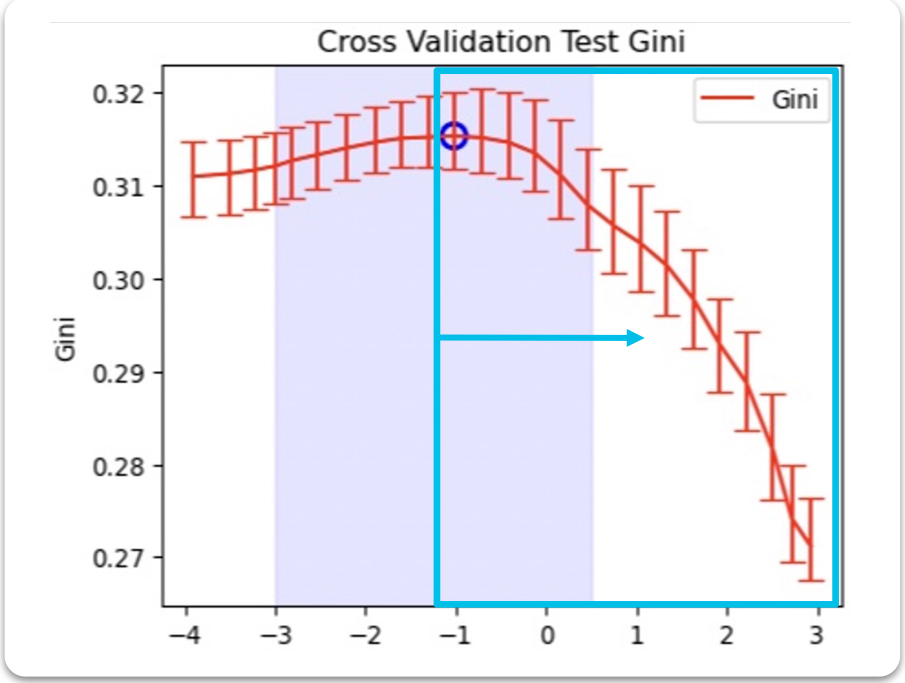
Excessive use of monotonic constraints to force desired behavior instead of real behavior.

- OK: Monotonic constraints applied on a few logical variables when data is thin or volatile.
- Watch: Monotonic constraints applied on many variables in a credible model.

In many cases, adjustments to the data to remove outliers can avoid monotonic constraints

Review of the Smoothness Parameter

Increased smoothness collapses coefficients to the prior assumption



Q: Is the selected model at the optimal smoothness or smoother?

Q: If the selected model is less smooth (more reactive to the data) than the optimal model, why was this smoothness selected?

Models that are smoother than the optimal model contain factors closer to a 1.0 factor or the complement of credibility.

This is best explained through **Lasso Credibility**.

Upcoming Monograph: Lasso Credibility

Implementing a Better Complement of Credibility

The offset can be used to set prior assumptions on all coefficients

$$
\begin{aligned}
\text{Prediction} &= \exp\left(\beta_0 + \text{offset} + \beta_1 X_1 + \beta_2 X_2\right) \\
&= \exp\left(\beta_0 + \beta_{1\ \text{offset}} X_1 + \beta_{2\ \text{offset}} X_2 + \beta_1 X_1 + \beta_2 X_2\right) \\
&= \exp\left(\beta_0 + \left(\beta_{1\ \text{offset}} + \beta_1\right) X_1 + \left(\beta_{2\ \text{offset}} + \beta_2\right) X_2\right)
\end{aligned}
$$

$$\text{Unpenalized coefficient} = \beta_{n,\text{offset}} + \beta_n = \beta_{n,\text{glm}}$$

$$\text{Fully penalized coefficient} = \beta_{n,\text{offset}} + \beta_n = \beta_{n,\text{offset}} + 0 = \beta_{n,\text{offset}}$$

$$\boxed{\text{Coefficient}_n = \left(\beta_{n,\text{offset}} + \beta_n\right) = Z \times \left(\beta_{n,\text{glm}}\right) + (1 - Z) \times \left(\beta_{n,\text{offset}}\right)}$$

$$\boxed{\beta_{n,\text{offset}} \leq \beta_{n,\text{offset}} + \beta_{n,\text{lasso}} \leq \beta_{n,\text{glm}}}$$

Calculating Indicated Relativities from a GLM

Creating rating tables from a GLM

β_1	is 1 without fire extinguisher, 0 with
β_2	is an integer representing age of home

$$\beta_0 = \log_e(100) \approx 4.605$$

$$\beta_1 = \log_e(1.2) \approx 0.182$$

$$\beta_2 = \log_e(1.01) \approx 0.01$$

Fire Extinguishers	Factor
No	$\exp(0.182 \times 1) = 1.200$
Yes	$\exp(0.182 \times 0) = 1.000$

Age of Home	Factor
0	$\exp(0.01 \times 0) = 1.000$
1	$\exp(0.01 \times 1) = 1.010$
2	$\exp(0.01 \times 2) = 1.020$
3	$\exp(0.01 \times 3) = 1.030$
4	$\exp(0.01 \times 4) = 1.041$

$$\begin{aligned} \text{Pure Premium} &= \exp(\beta_0) \times \beta_1 X_1 \times \beta_2 X_2 \\ &= \exp(4.6057) \times \exp(0.182 \times X_1) \times \exp(0.01 \times X_2) \end{aligned}$$

Calculating Indicated Relativities in Lasso Credibility

Creating rating tables from lasso credibility

$$\beta_{1\ \text{offset}} = 0.182$$
$$\beta_{2\ \text{offset}} = 0.01$$

$$\beta_0 = 4.6057$$
$$\beta_1 = -0.087$$
$$\beta_2 = 0.01$$

Fire Extinguishers	Factor
No	$\exp((0.182 - 0.087) \times 1) = 1.100$
Yes	$\exp((0.182 - 0.087) \times 0) = 1.000$

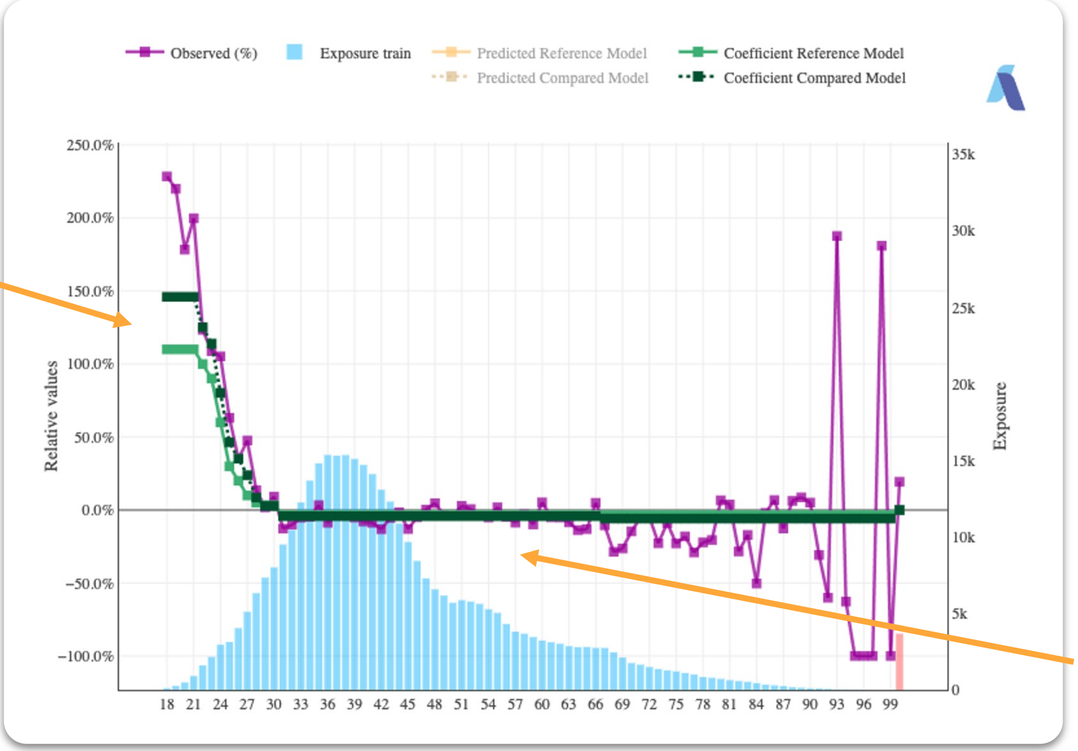
Age of Home	Factor
0	$\exp((0.01 + 0.01) \times 0) = 1.000$
1	$\exp((0.01 + 0.01) \times 1) = 1.020$
2	$\exp((0.01 + 0.01) \times 2) = 1.040$
3	$\exp((0.01 + 0.01) \times 3) = 1.061$
4	$\exp((0.01 + 0.01) \times 4) = 1.082$

$$\begin{aligned}
\text{Prediction} &= \exp\left(\beta_0 + \text{offset} + \beta_1 X_1 + \beta_2 X_2\right) \\
&= \exp\left(\beta_0 + \beta_{1\ \text{offset}} X_1 + \beta_{2\ \text{offset}} X_2 + \beta_1 X_1 + \beta_2 X_2\right) \\
&= \exp\left(\beta_0 + \left(\beta_{1\ \text{offset}} + \beta_1\right) X_1 + \left(\beta_{2\ \text{offset}} + \beta_2\right) X_2\right)
\end{aligned}$$

Driver Age Relativity Plot

Complement = light green | Complement + modeled coefficient = dark green

Significant difference identified by the model.



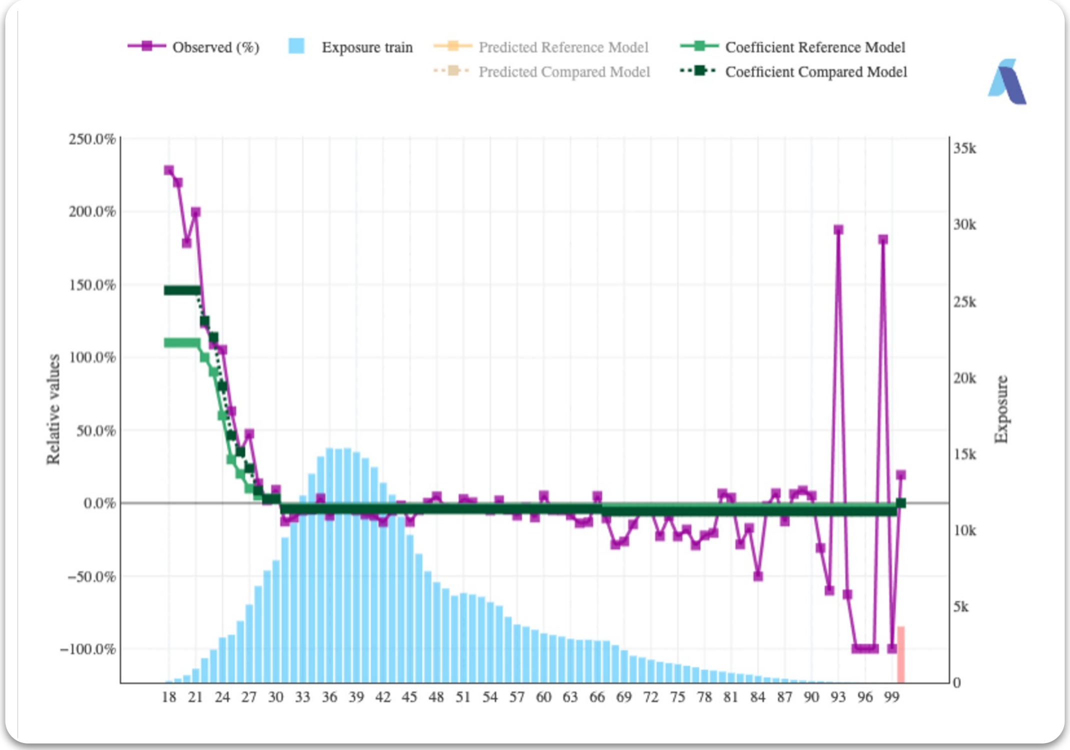
Model is not significantly different – no Change

Driver Age Modeled with a Complement of Credibility

Complement = light green | Complement + modeled coefficient = dark green

As the penalty increases, the modeled coefficient moves to zero.

When the modeled coefficient is zero, the complement has full credibility.



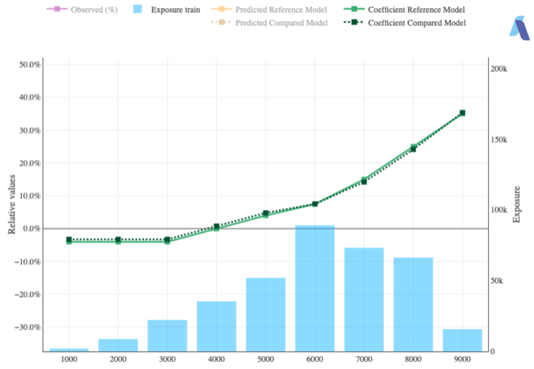
Lasso Credibility – Anticipated Use Case for Model Refit

Simple to build, simple to review, a multivariate credibility procedure

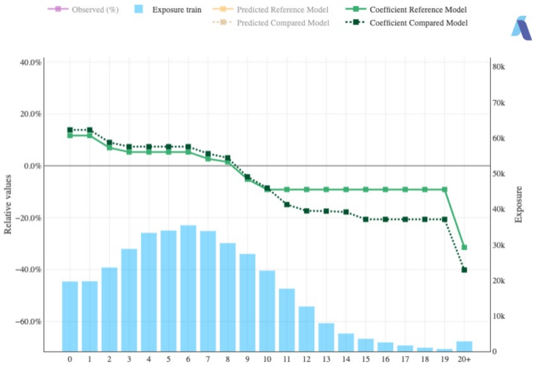
Assumptions:

- Current rating factors are selected as the complement of credibility for the model.
- These rating characteristics are intuitive and not subject to additional scrutiny.
- Underlying data does not have material issues or outliers.

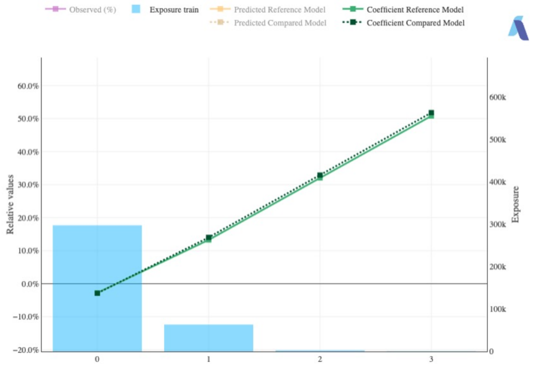
Mileage – No Credible Change



Vehicle Age –Credible Change



Accident Count – No Credible Change



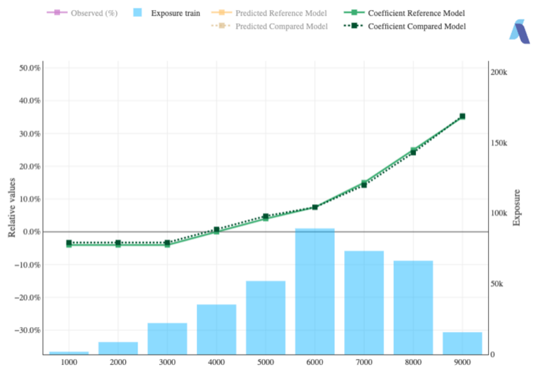
Lasso Credibility – Anticipated Use Case for Model Refit

Simple to build, simple to review, a multivariate credibility procedure

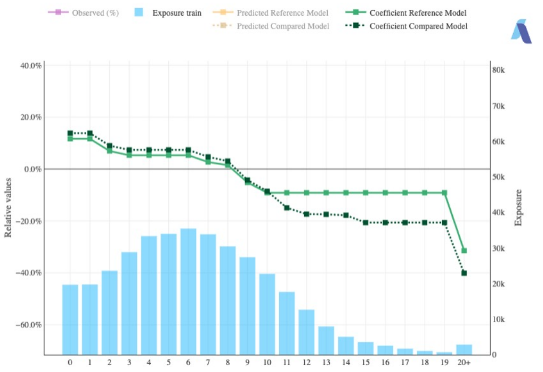
Review Outcomes:

1. Model is too reactive. Smoothness needs to be increased.
2. Model is technically appropriate. Review focuses on post-modeling decisions and selections.

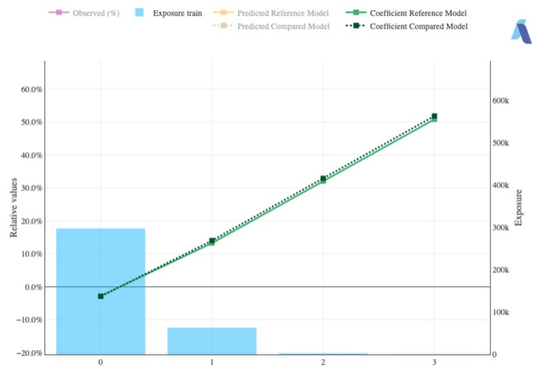
Mileage – No Credible Change



Vehicle Age –Credible Change



Accident Count – No Credible Change



Lasso Credibility – Anticipated Use Case for Model Refit

Simple to build, simple to review, a multivariate credibility procedure

Goals of Lasso Credibility Adoption for Insurers:

- **Increased speed and accuracy of modeling** by incorporating a complement of credibility.
- **Stability of rates** by not fitting a GLM from scratch for each refit.
- Clear distinction between technical and actuarial selections made by modelers.
- **Ease of technical review** due to simplicity of model structure and foundation in credibility.

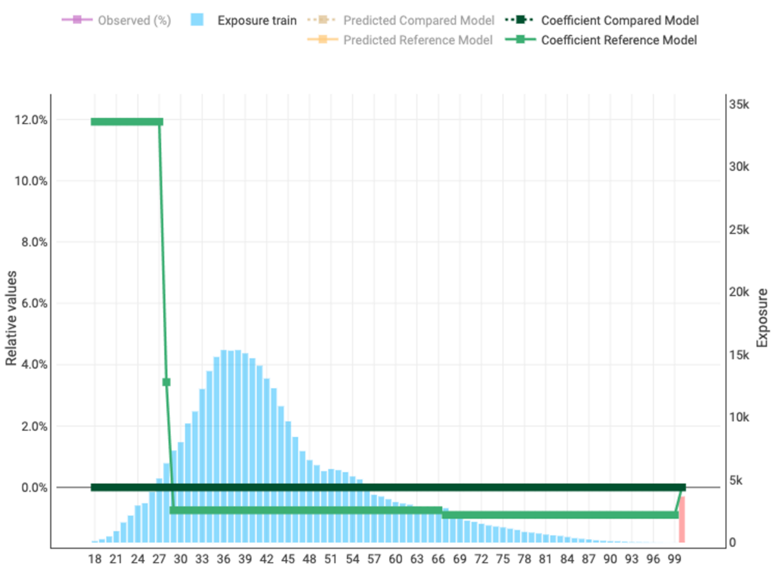
Similar to the review process for standard credibility procedures, typical model review for lasso credibility will focus **highly on industry expertise** and **lightly on technical expertise**.

New Items for Review in Lasso Credibility

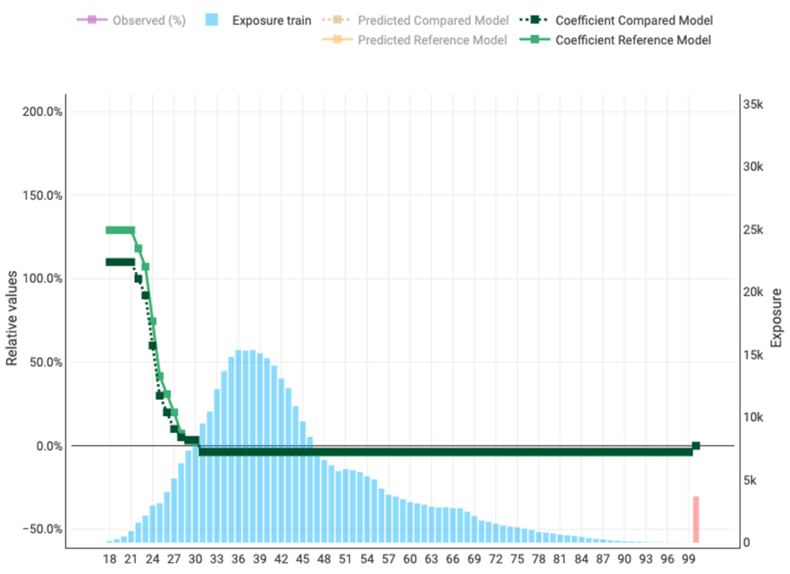
Relativity Plots

Factors only

Traditional Univariate Plot



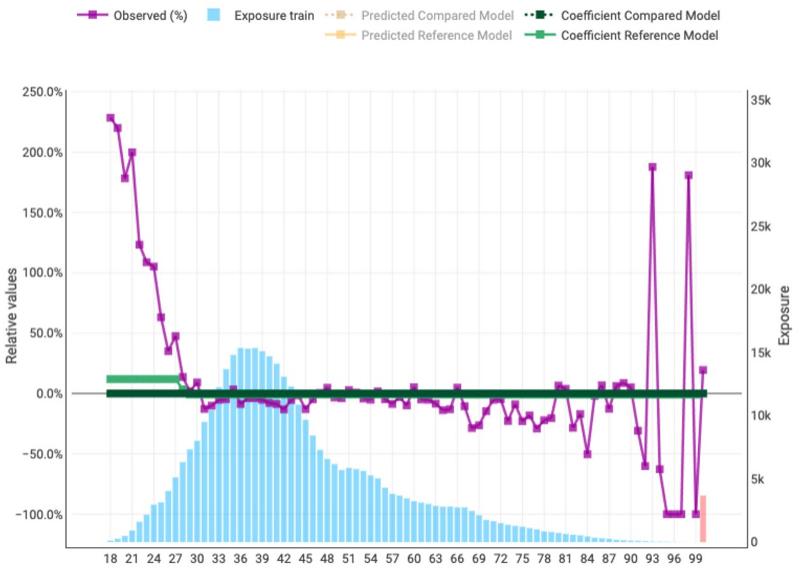
Recommended Relativity Plot



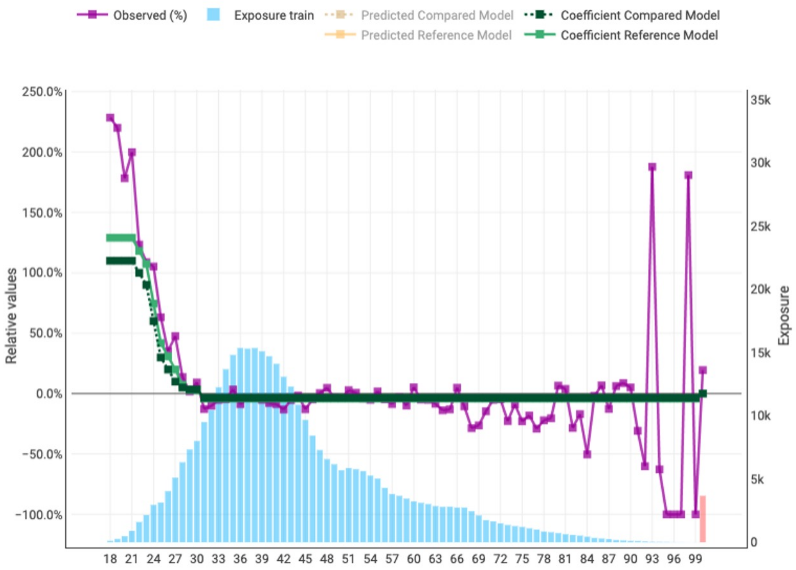
Relativity Plots

Factors and observed

Traditional Univariate Plot



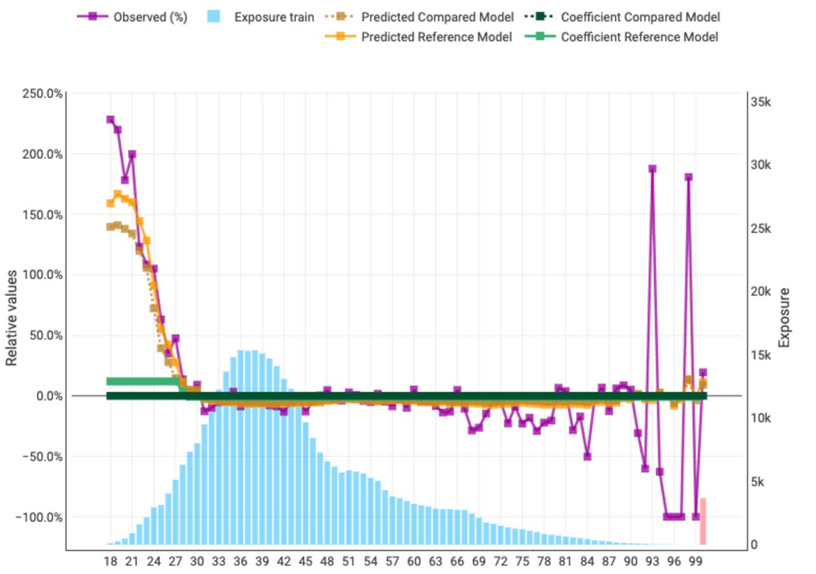
Recommended Relativity Plot



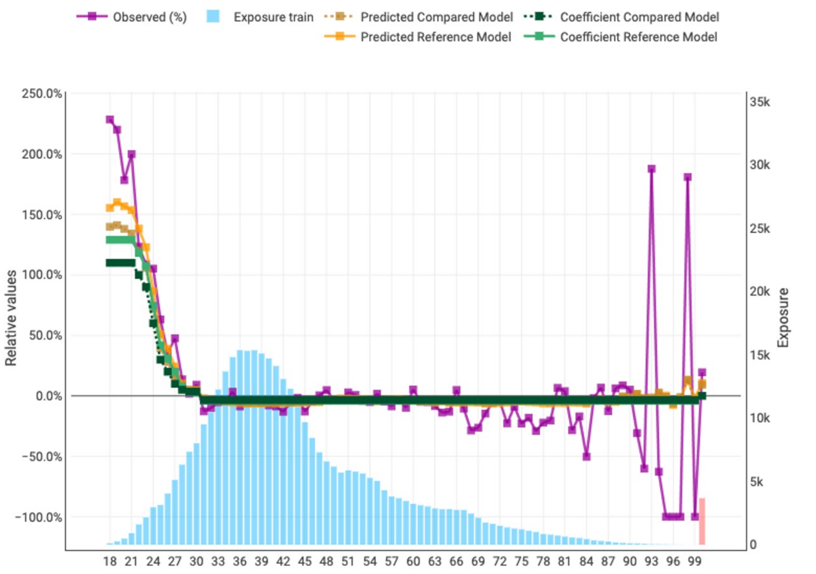
Relativity Plots

Factors, predicted, and observed

Traditional Univariate Plot



Recommended Relativity Plot



Review of the Complement of Credibility

Upcoming monograph will include discussion of this review

An ideal complement of credibility comes from a single source:

- Current implemented rating plan.
- Bureau approved rating plan.

Watch Items:

- A complement of credibility using variables from different sources or rating plans.
- A complement of credibility is used for some variables and not others where it exists.
- Indicated coefficients significantly deviate from a complement where there is little data.
- Manually selected deviations from a selected complement of credibility.

Conclusion

Conclusions

Upcoming papers will provide additional resources

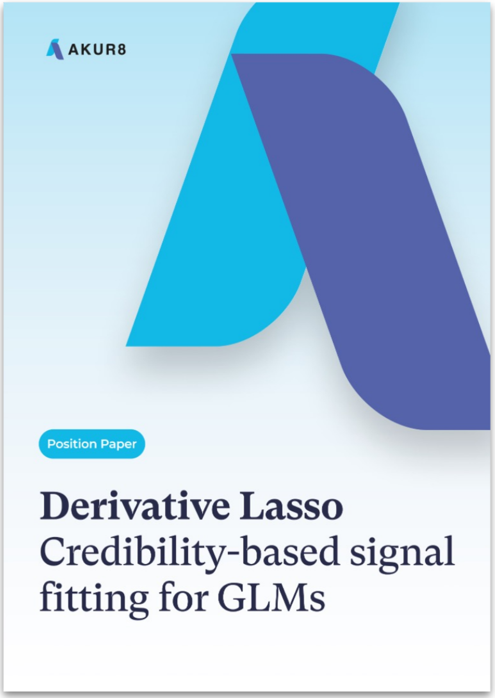
Derivative lasso is an application of lasso penalization specifically designed to create fast, transparent, and actuarially appropriate pricing models.

Lasso credibility is an application of the offset and lasso penalization that, when combined with derivative lasso methodology, is a transparent multivariate credibility procedure.

Derivative Lasso: Credibility-Based Signal Fitting for GLMs will be published by Akur8 later this year. This will thoroughly discuss the derivative lasso modeling approach and its genesis.

Penalized Regression and Lasso Credibility monograph is in review with the CAS monograph committee. This will provide the statistical background for lasso credibility along with additional guidance for modeling and review.

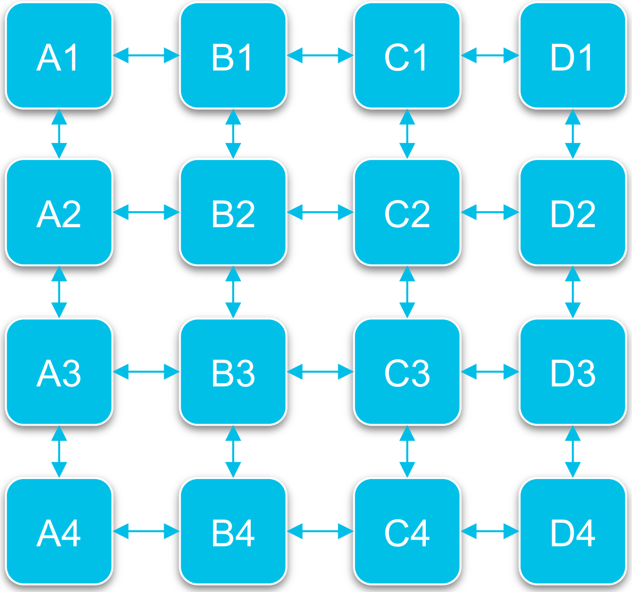
For further discussion: thomas.holmes@akur8.com



Appendix

One Tuning Parameter: Smoothness

Not mgcv smoothness



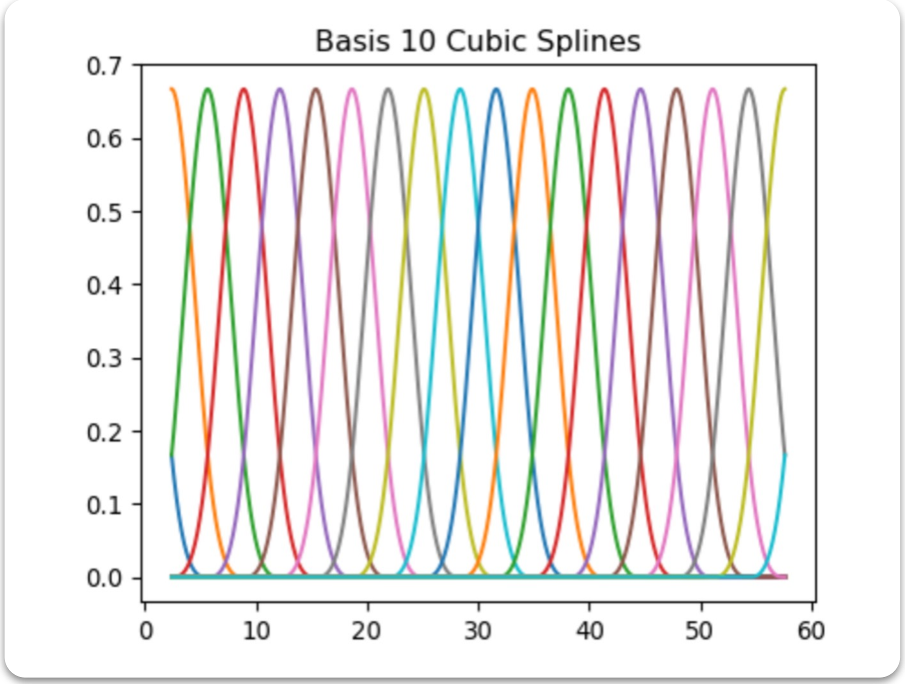
The derivative lasso penalty generalizes for n-dimensional interactions.

In two dimensions, derivative lasso creates a highly interpretable interaction through the application of a “net” of penalty terms.

A proper smoothness term will create an intuitive smooth surface of factors.

MGCV GAM: Spline Basis Function

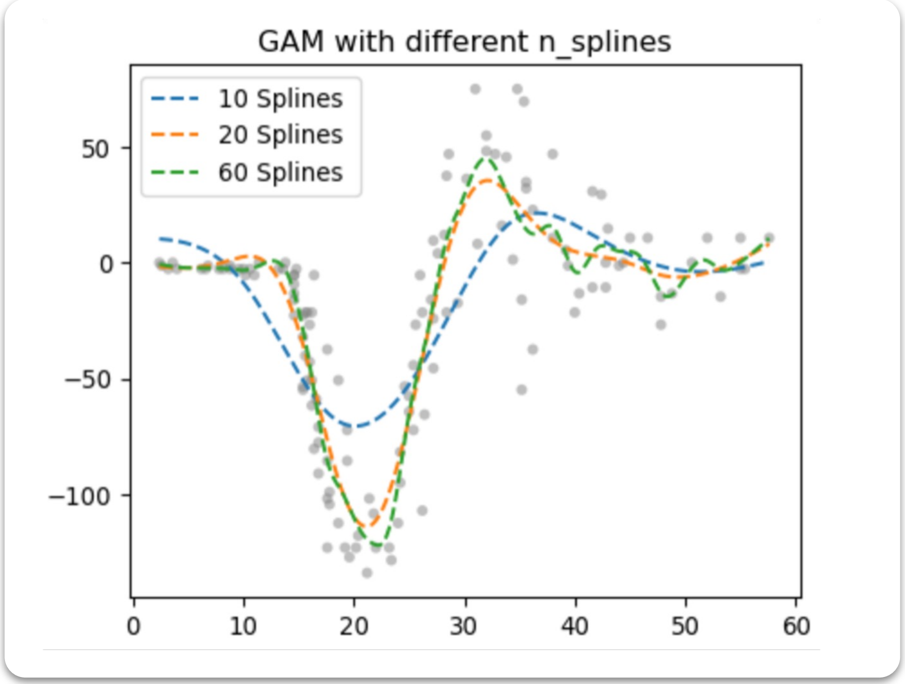
Complex curve-fitting



GAM as defined by the MGCV package can use a series of cubic splines to fit complex curves.

MGCV GAM: Tuning the Number of Splines

Complex curve fitting



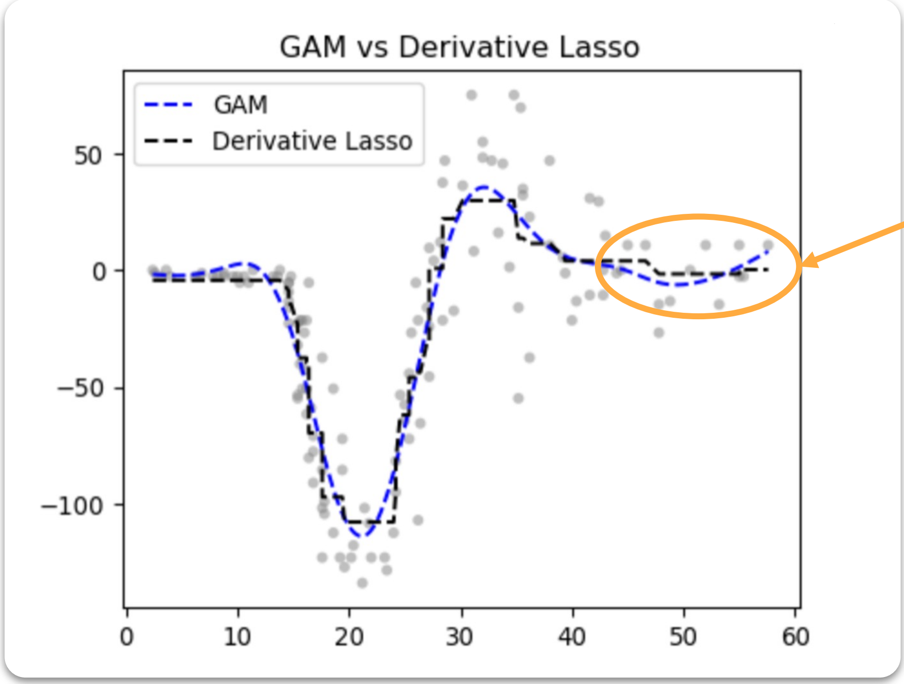
Tuning the number of splines is done judgmentally on a variable-by-variable basis.

This tuning is highly manual and iterative.

The appropriate penalty parameter will change based on the selected number of splines.

Tuned MGCV GAM vs. Tuned Derivative Lasso Model

Derivative lasso has actuarial benefits over curve-based basis functions



Derivative Lasso is able to fit this complex curve with a single parameter.

A well-tuned cubic-spline GAM is still subject to “whiplash” due to the curve-based nature of the basis function.