March 13–15, 2023
San Diego, CA
Loews Coronado Bay Resort

Ratemaking, Product and Modeling Seminar

AKUR8

Practical Impact of Credibility Models

2023

Practical Impact of Credibility Models

Summary

Introductory Poll

Introduction to Penalization

Four Practical Scenarios

Practical Applications of Penalized Regression

Presenter Bios

Thomas Holmes is the Head of US Actuarial Data Science at Akur8. In this role, he works directly with clients providing modeling insights and best practices with Akur8's software. Additionally, he works with Akur8’s product development team to improve the software's functionality for the US region. Prior to Akur8, Thomas has worked in both traditional ratemaking and modeling roles. Thomas is a Fellow of the CAS.

Jake Falandays is the Director of Actuarial Pricing at Hippo, where he designs and applies pricing models and advances Hippo’s data strategy serving proactive home protection.

Jake is a Fellow of the CAS and has 10+ years of experience in capital modeling, pricing, and analytics roles at various global carriers.

James Riley is an Assistant Actuary at Allstate Insurance Company. In this role, he works on developing actuarial pricing models across multiple lines of business. James has a B.S. in Actuarial Science and Applied Statistics from Purdue University and is a Fellow of the CAS.

Background Knowledge Poll

CAS App Polling

What area do you currently work in?

- Modeling
- Pricing
- Reserving
- Regulation
- Other

Background Knowledge Poll

CAS App Polling

What is your familiarity with GLM or penalized GLM modeling?

- I have built pricing models using GLMs
- I have evaluated GLMs built by others
- I have a theoretical knowledge of GLMs
- I have little knowledge about GLMs

Background Knowledge Poll

CAS App Polling

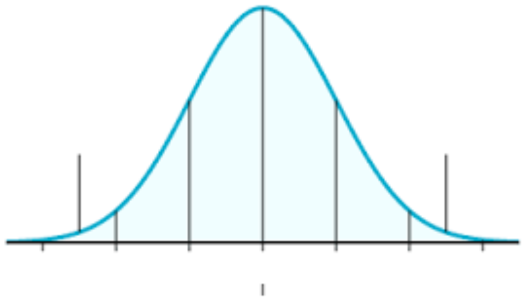
What is your familiarity with penalized regression? (Lasso, Ridge, Elastic Net)

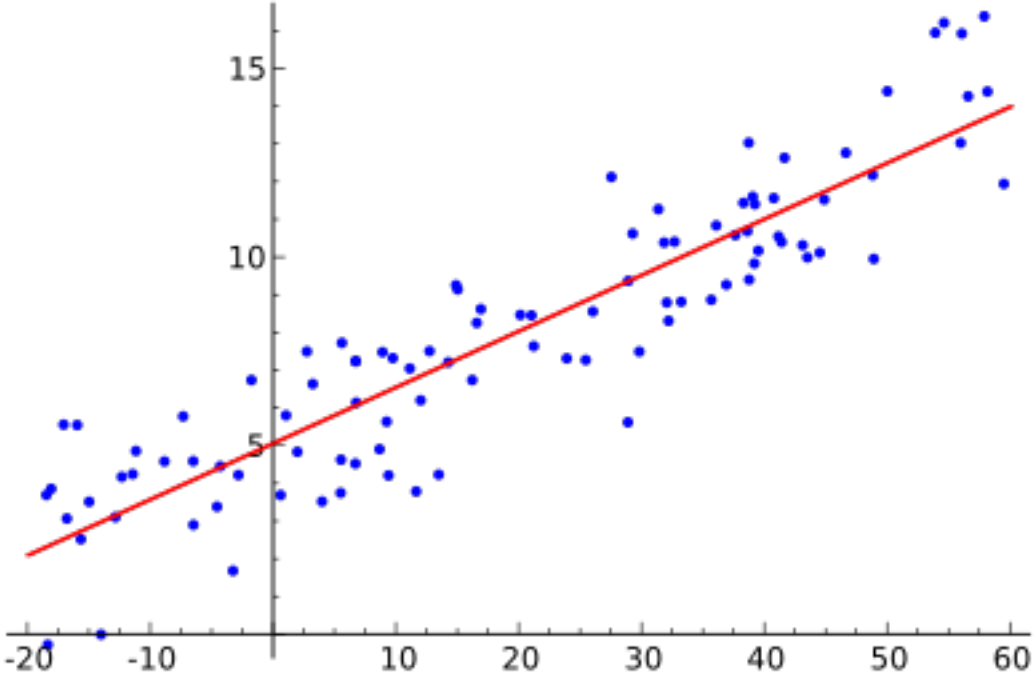
- I have built pricing models using penalized regression
- I have evaluated penalized regression models built by others
- I have a theoretical knowledge of penalized regression
- I have little knowledge about penalized regression

Linear Regression – No Formulas!

(Not a GLM!)

In Linear Regression, we are basically placing the line in such a way that it **minimizes the error** between the estimate and the datapoints.

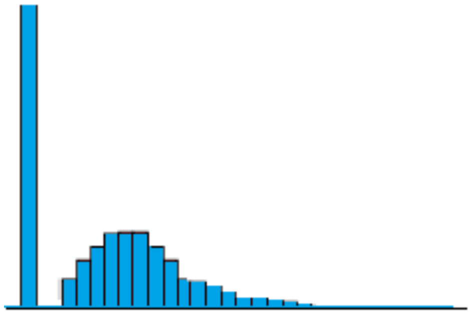


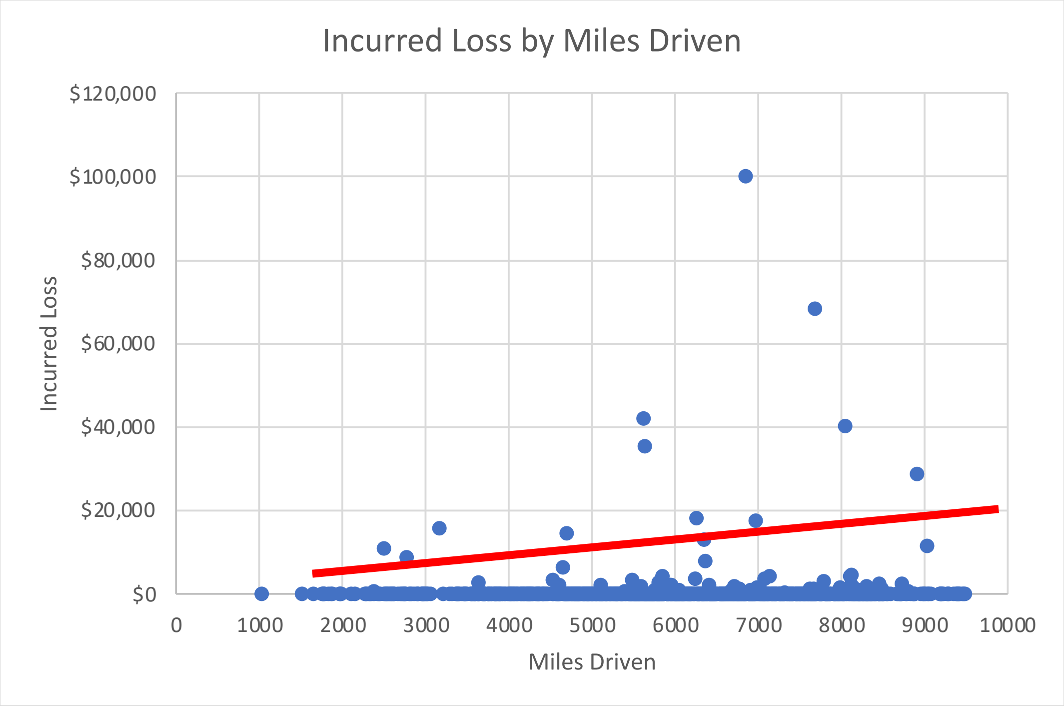


GLMs – No formulas!

Generalized Linear Models

In a GLM, we are placing the line so that it **maximizes the likelihood** of each event given our distribution and estimate.





*Not an actual model fit

The estimate of a GLM is Based on the Selected Distribution

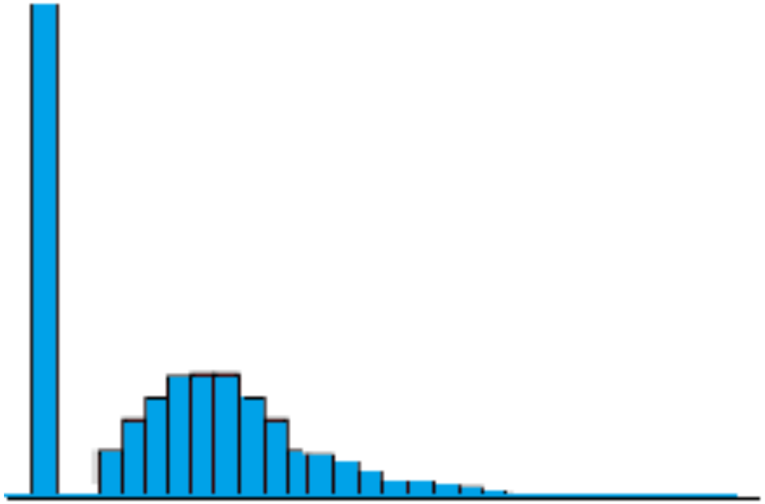
Ok, now we have formulas. Remember Exam P?

A model with no variables will contain only an **intercept** representing the **mean of the distribution**.

Y = Intercept

Adding coefficients **adjusts the mean** of the expected loss distribution.

Y = Intercept +Bx

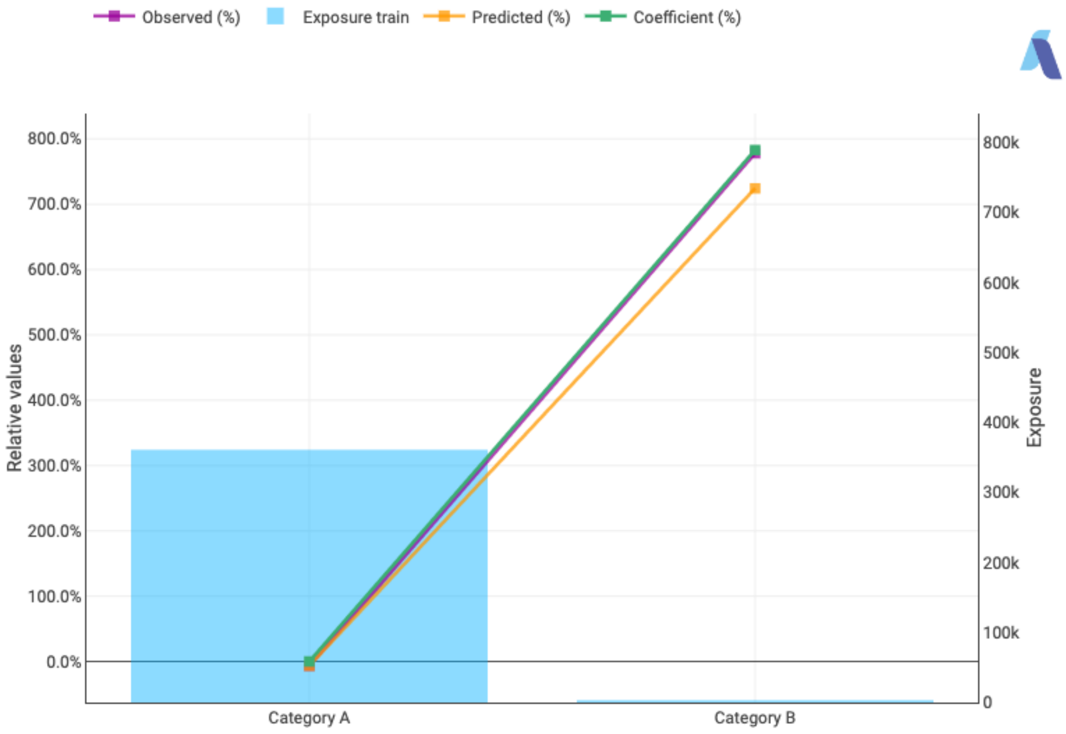


Penalizing a One-Variable Model

Penalization is Credibility

GLM:
Y = Intercept + Bx

GLMs treat the data with full credibility

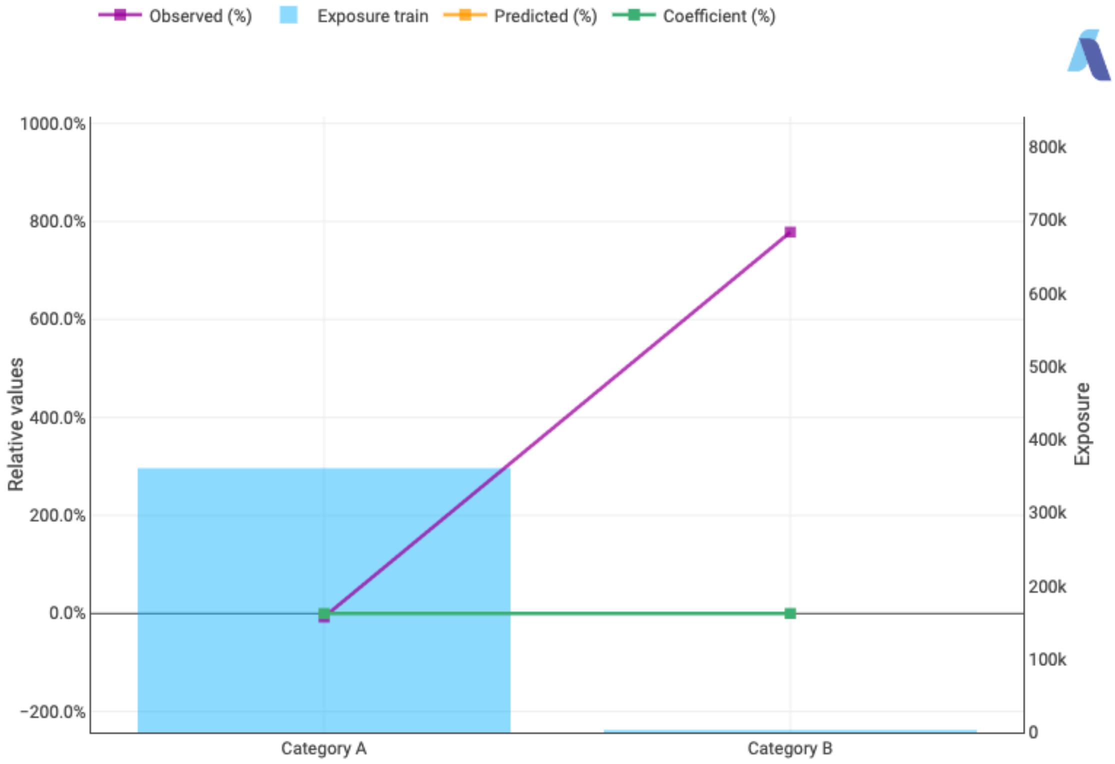


Penalizing a One-Variable Model

Penalization is Credibility

(Fully Credible) GLM:
Y = Intercept + Bx

Fully Penalized GLM:
Y = Intercept + 0x



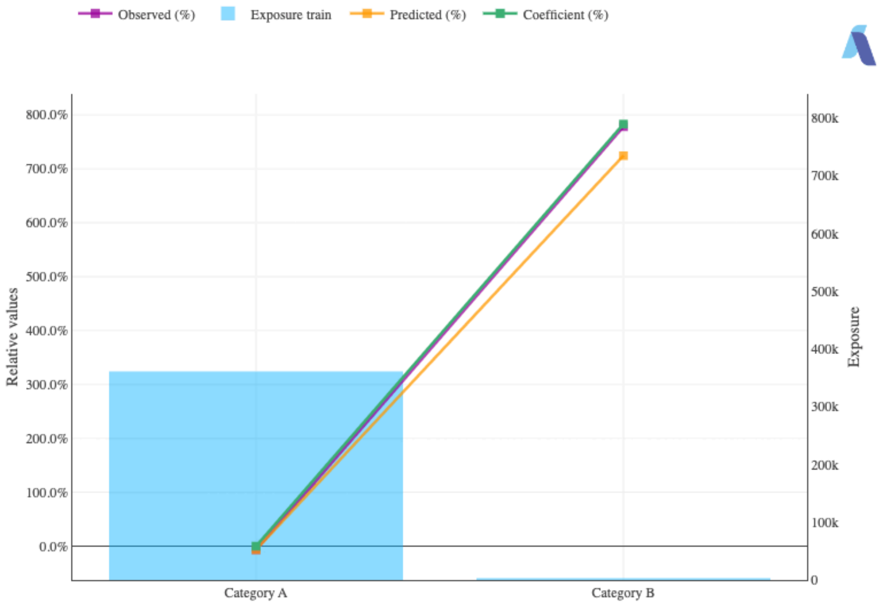
Penalizing a One-Variable Model

Penalization is Credibility

(Fully Credible) GLM:
Y = Intercept + Bx

Fully Penalized GLM:
Y = Intercept + 0x

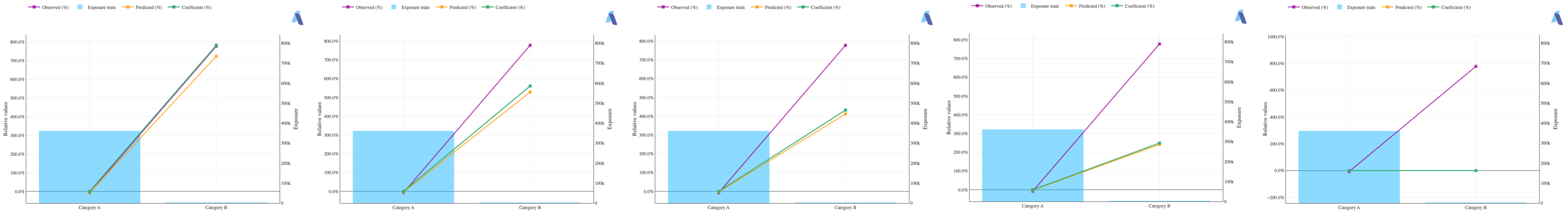
Partially Penalized GLM:
Y = Intercept + B*x
B > B* > 0



When a Modeler Selects a Penalty Parameter, they Apply Credibility

There are tools to assist the selection of a penalty parameter

$$0 \;\ldots\; small\ penalty \;\ldots\; \lambda \;\ldots\; large\ penalty \;\ldots\; \infty$$



no credibility … … low credibility … … **Overall Average** … … high credibility … … **full credibility**

full credibility … … high credibility … … **Modeling Data** … … low credibility … … **no credibility**

Literature on Penalized Regression and Credibility

Multiple papers exist on this subject – and more will exist soon!

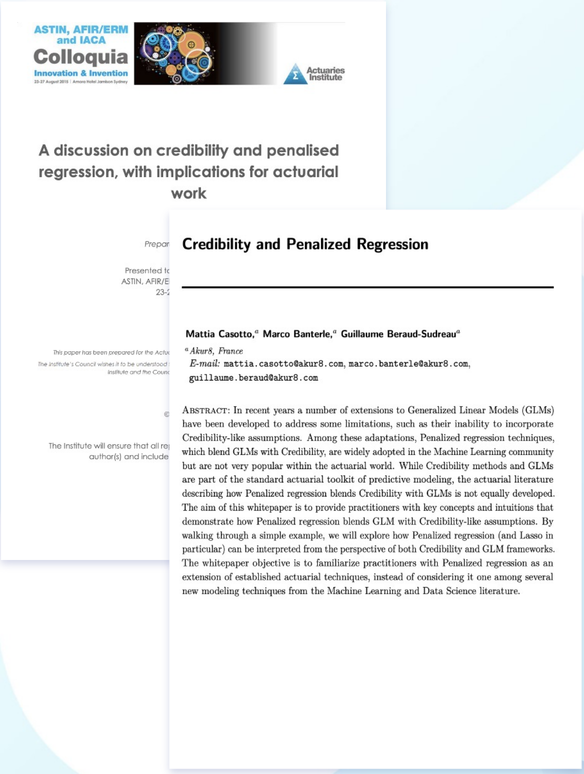
For more information on this link, please reference the following papers:

- M.Casotto et al. [“Credibility and Penalized Regression”]() (2022)
  (this topic was previously presented at CAS events)
- Fry, Taylor. ["A discussion on credibility and penalised regression, with implications for actuarial work"]() (2015)

Upcoming Call for Reviewers on a CAS Monograph:

- *Contact Brandon Smith if interested: brandon.smith@markel.com*

(Don’t reach out to me – reviewers should be anonymous)



ASTIN, AFIR/ERM and IACA Colloquia
Innovation & Invention
Actuaries Institute

A discussion on credibility and penalised regression, with implications for actuarial work

Credibility and Penalized Regression

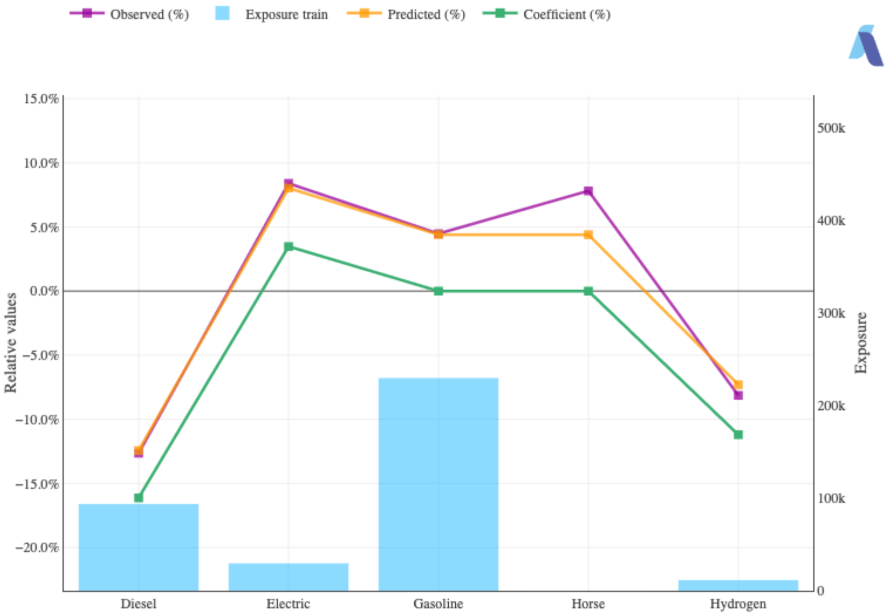
Mattia Casotto,[a] Marco Banterle,[a] Guillaume Beraud-Sudreau[a]
[a] Akur8, France
E-mail: mattia.casotto@akur8.com, marco.banterle@akur8.com, guillaume.beraud@akur8.com

Abstract: In recent years a number of extensions to Generalized Linear Models (GLMs) have been developed to address some limitations, such as their inability to incorporate Credibility-like assumptions. Among these adaptations, Penalized regression techniques, which blend GLMs with Credibility, are widely adopted in the Machine Learning community but are not very popular within the actuarial world. While Credibility methods and GLMs are part of the standard actuarial toolkit of predictive modeling, the actuarial literature describing how Penalized regression blends Credibility with GLMs is not equally developed. The aim of this whitepaper is to provide practitioners with key concepts and intuitions that demonstrate how Penalized regression blends GLM with Credibility-like assumptions. By walking through a simple example, we will explore how Penalized regression (and Lasso in particular) can be interpreted from the perspective of both Credibility and GLM frameworks. The whitepaper objective is to familiarize practitioners with Penalized regression as an extension of established actuarial techniques, instead of considering it one among several new modeling techniques from the Machine Learning and Data Science literature.

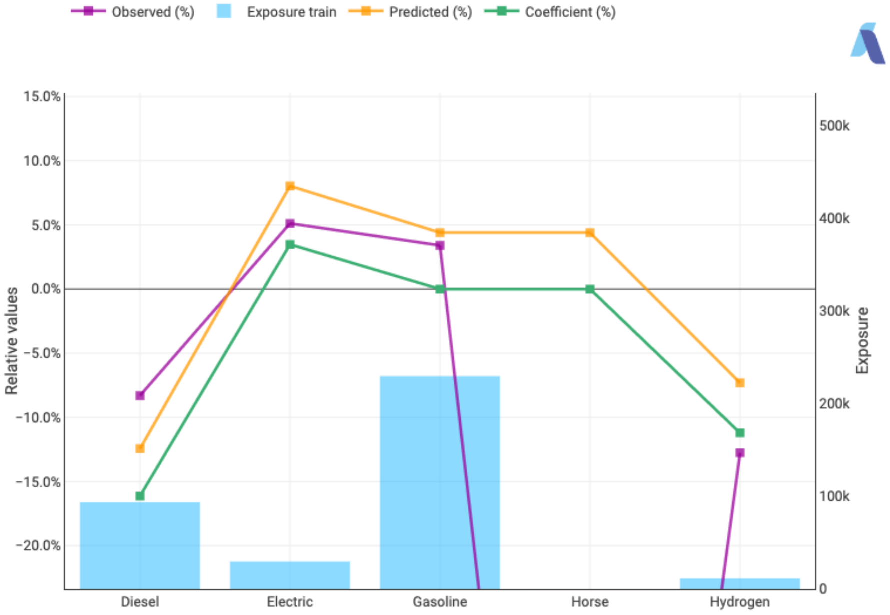
Penalized Regression – Ordinal/Continuous Variable Vehicle Age

Small Penalty

Training Performance: **Good**



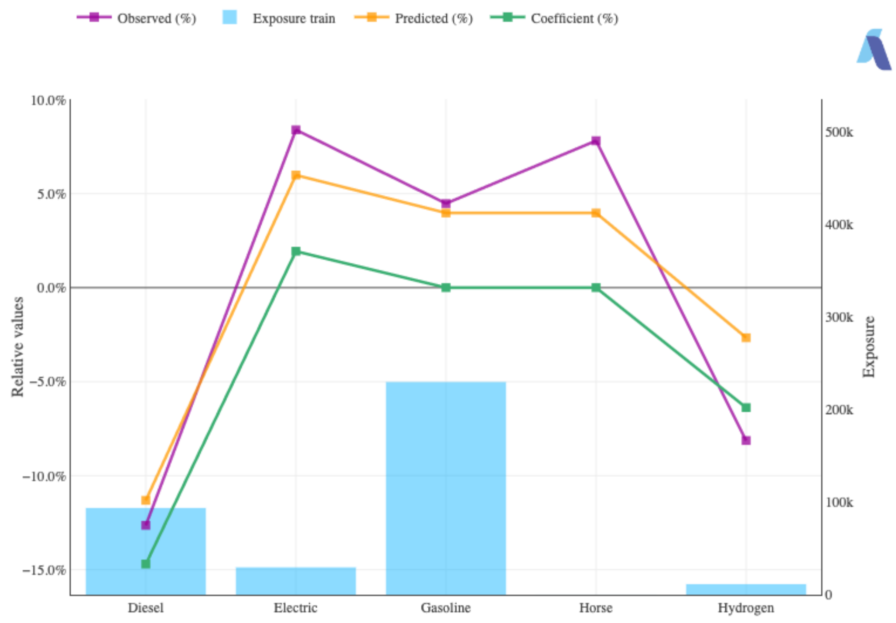
Holdout Performance: **OK**



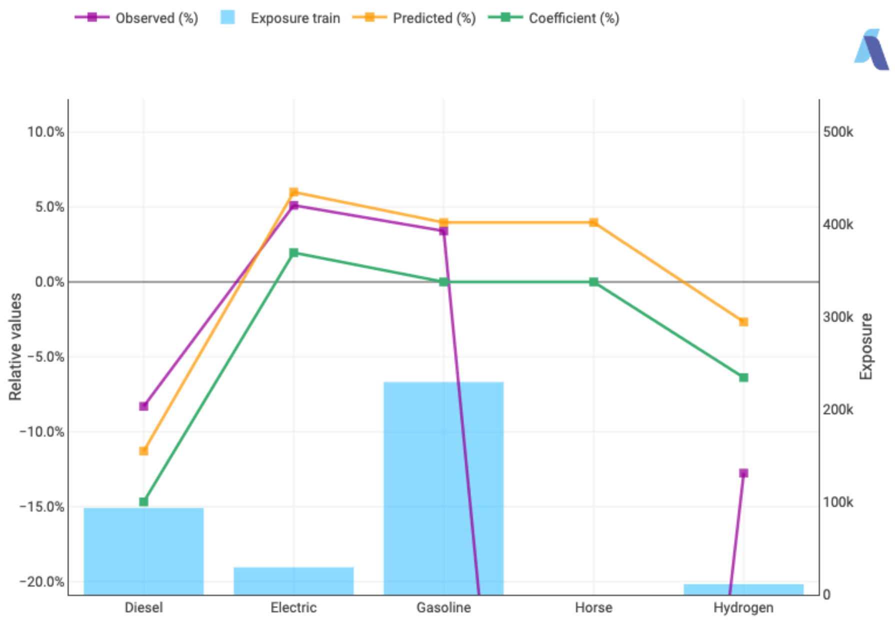
Penalized Regression – Ordinal/Continuous Variable Vehicle Age

Medium/Appropriate Penalty

Training Performance: OK



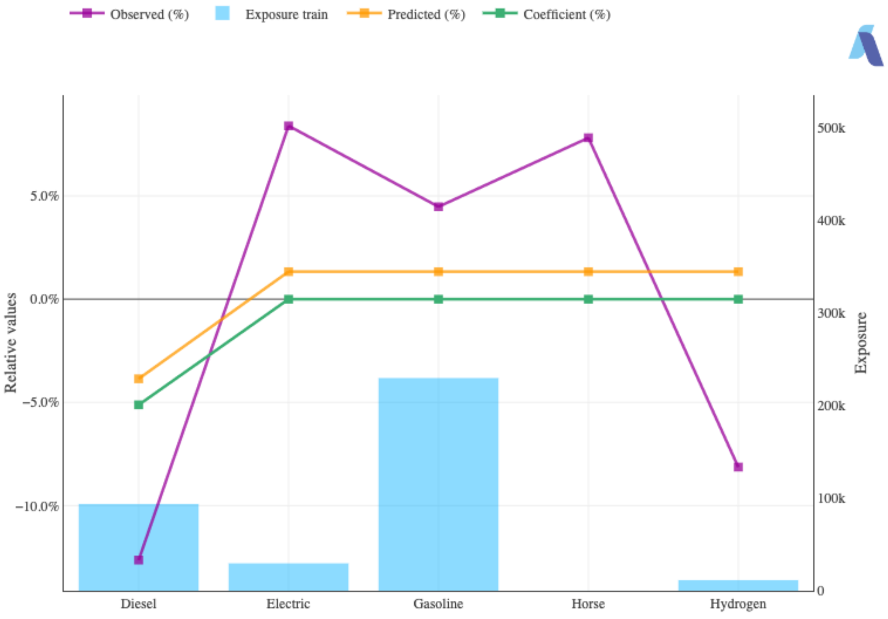
Holdout Performance: Good



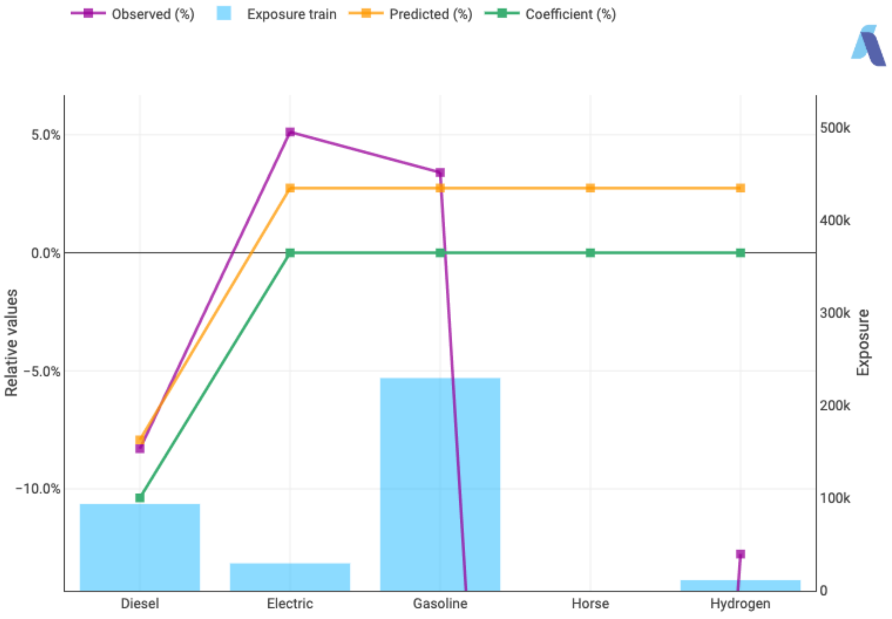
Penalized Regression – Ordinal/Continuous Variable Vehicle Age

Large Penalty

Training Performance: **Bad**



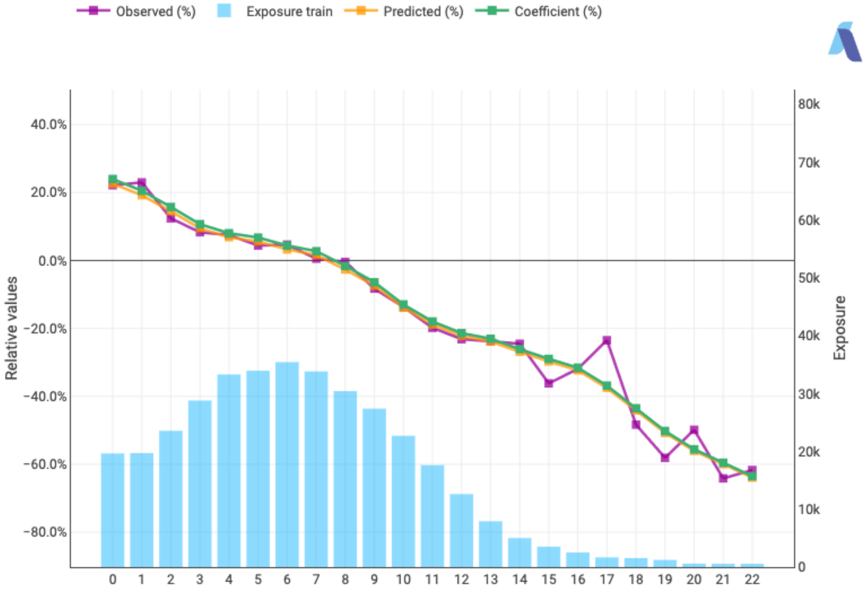
Holdout Performance: **Bad**



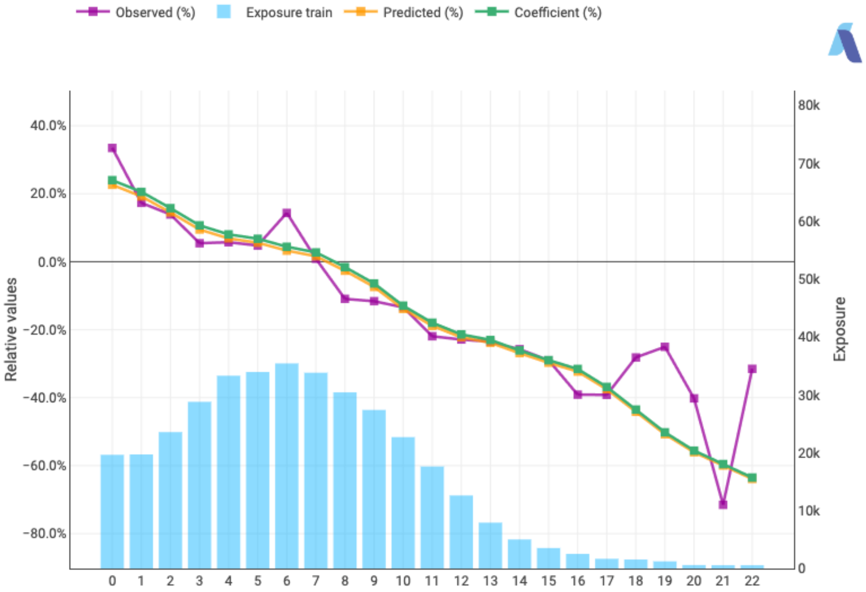
Penalized Regression – Ordinal/Continuous Variable Vehicle Age

Small Penalty

Training Performance: Good



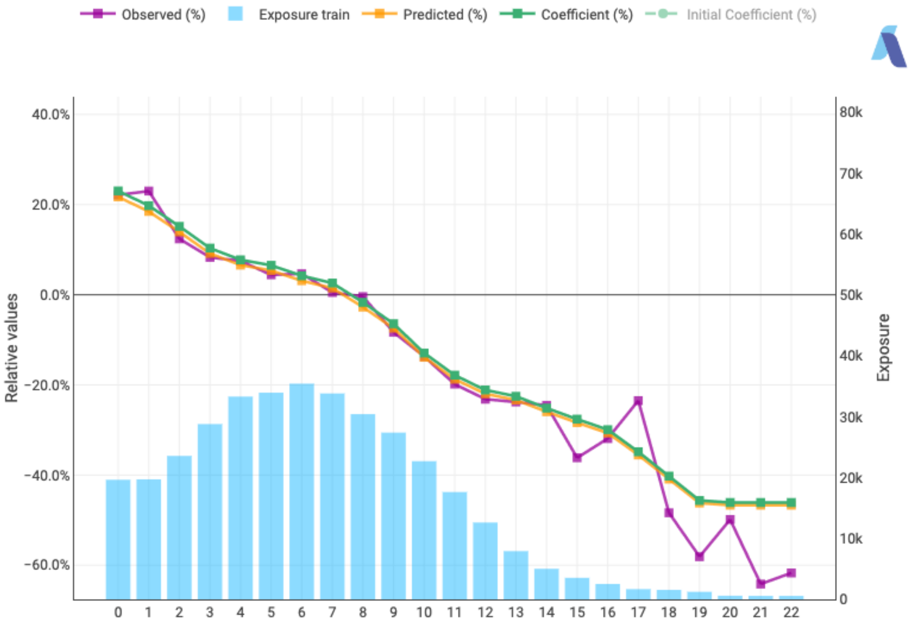
Holdout Performance: OK



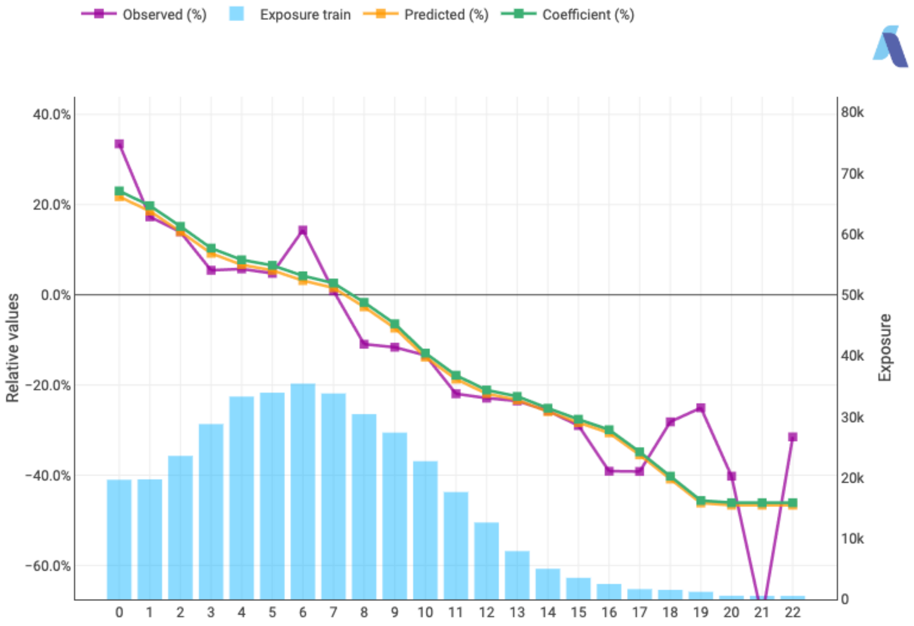
Penalized Regression – Ordinal/Continuous Variable Vehicle Age

Medium/Appropriate Penalty

Training Performance: **OK**



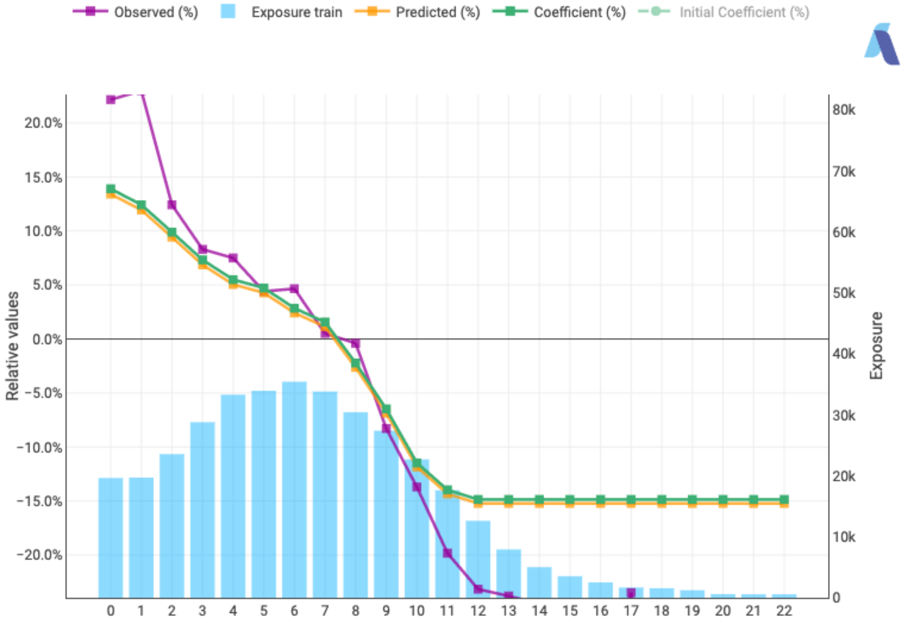
Holdout Performance: **Good**



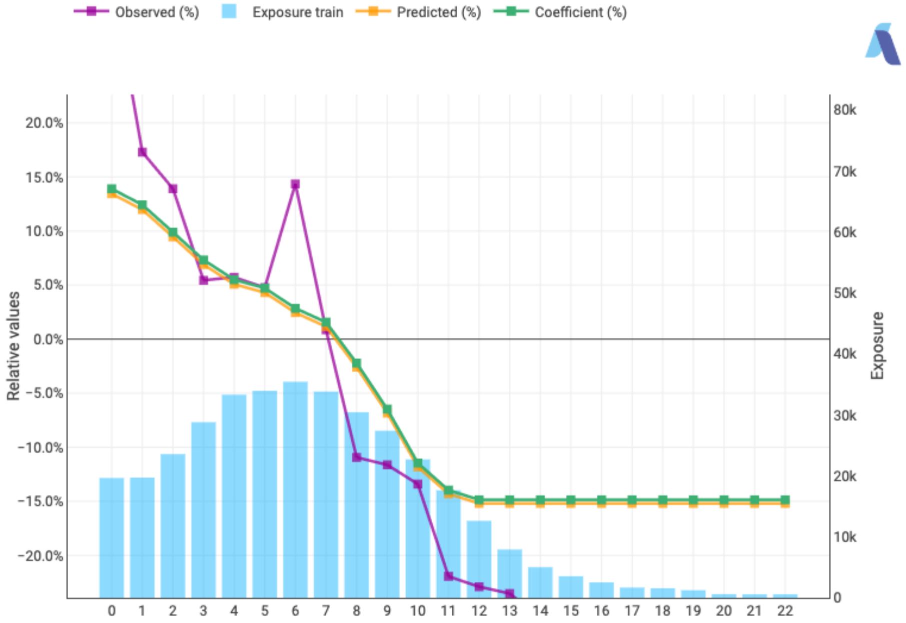
Penalized Regression – Ordinal/Continuous Variable Vehicle Age

Large Penalty

Training Performance: Bad

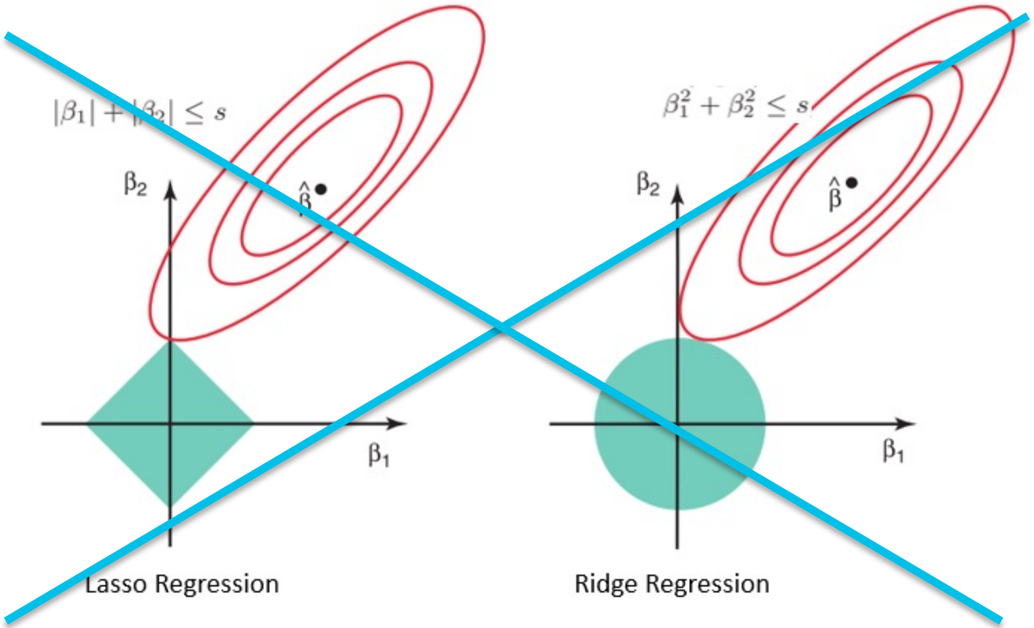


Holdout Performance: Bad



How to Think about Practical Penalized Regression

This visualization has never been helpful for me



Lasso:
Coefficients can be penalized to zero

Ridge:
Coefficients can **not** be penalized to zero

Lasso: *Similar* to Classical Credibility

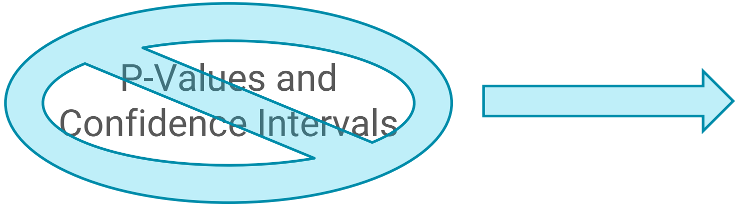
Ridge: *Similar* to Buhlmann Credibility*

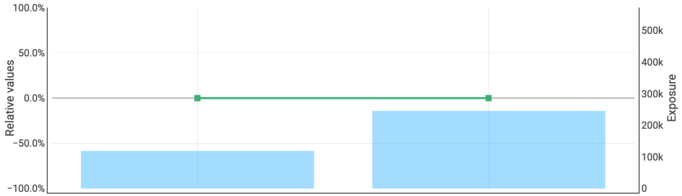
* Equivalent in a special case, see papers

Penalized Regression and Benefits to Model Review

Fair and uniform credibility treatment

- A properly tuned penalty parameter **automatically removes insignificant variables**





- **Credibility is applied uniformly** within the model

Credibility Treatment For Variable1 ⟷ Credibility Treatment For Variable2 ⟷ Credibility Treatment For Variable3 ⟷ [...] ⟷ Credibility Treatment For VariableN

- **Penalization applies credibility** to appropriately consider good or bad experience in smaller segments

Practical Scenarios

Small Data - Continuous

Practical Scenario – Modeling on a small growing book

Scenario set up for mock homeowners' book

The Set Up:

You're the actuary for a small, fast-growing homeowners' book

You're asked to build the company's first loss cost model, but you're worried about your thin experience – management's primary concern is that your model (and prices) are **stable** year over year

You expect from experience that age of home will be a major predictor, so you build a modeling data set with age of home and some control variables

Note: all data displayed is a mock data set

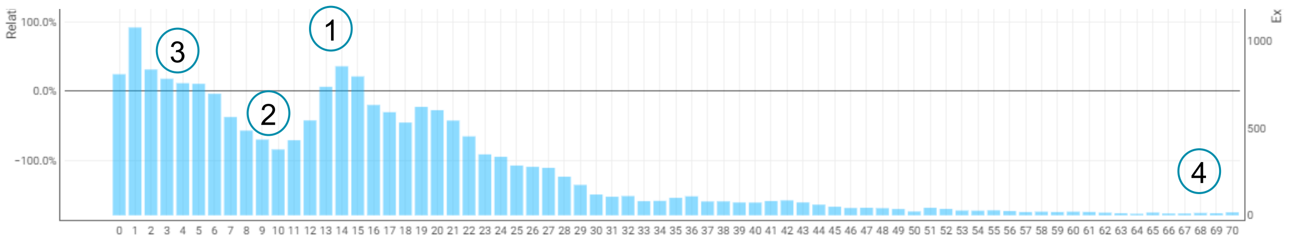
Modeling a continuous variable with peaks, valleys, and tails

Treatment for Numeric Variables

Unusual exposure distribution creates segments of varying credibility

1. Exposure peaks at height of new construction in US
2. Exposure low point when housing construction dipped following 2008-2009 recession
3. The book has plenty of newer homes
4. The book has few very old homes

Exposures by Age of Home

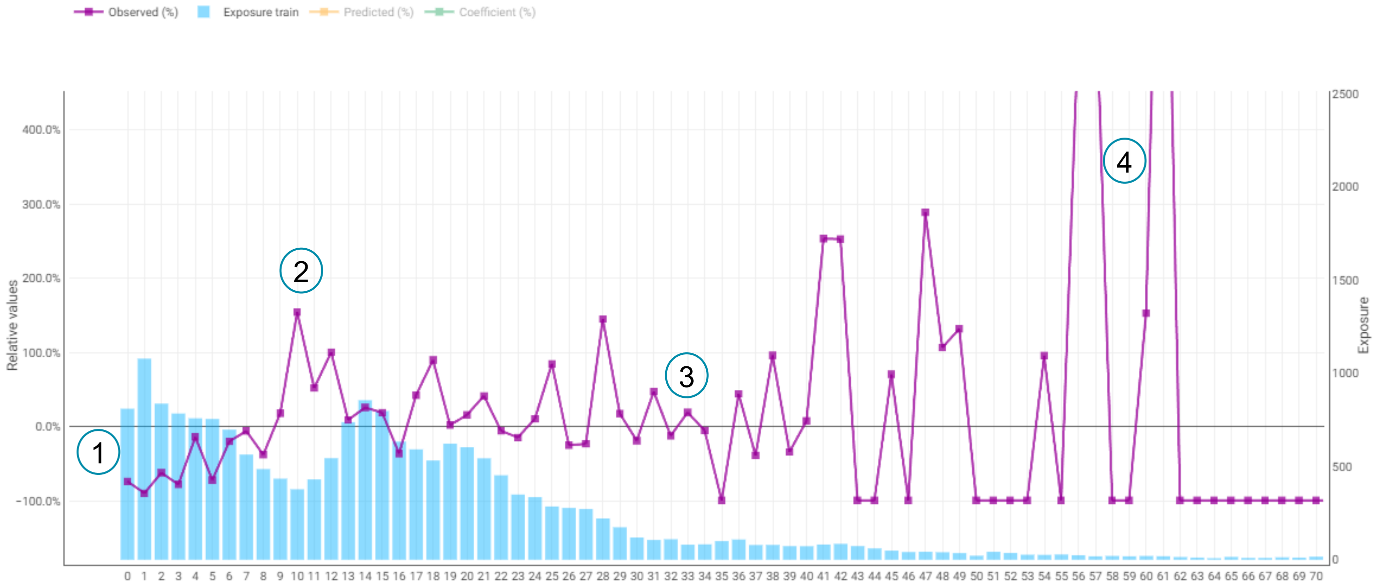


Modeling a continuous variable with peaks, valleys, and tails

Treatment for Numeric Variables

Open question: What do you, the modeler, see that might require special handling?

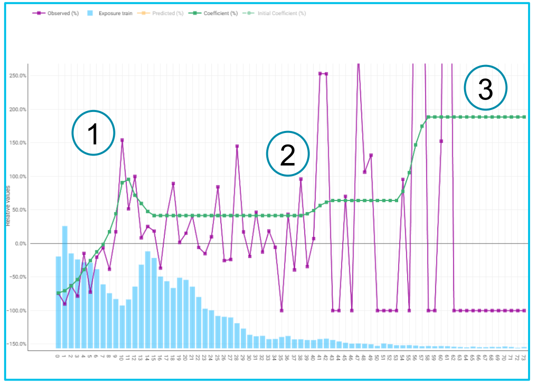
1. Newer homes have lower observed claim frequency
2. The observed climbs to a peak at age of home 10, **which coincides with the exposure low-point**
3. Noisy reversals in the middle
4. Some heightened experience in the older homes, but volatile due to thin exposures

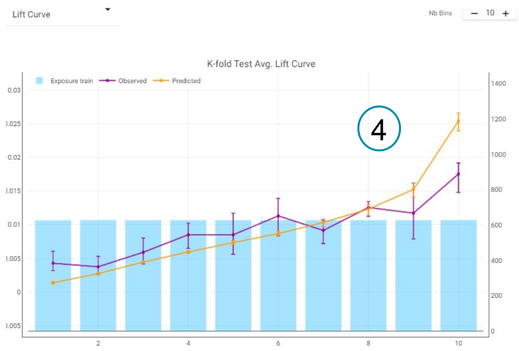


Modeling a continuous variable with peaks, valleys, and tails

A GLM-like, low-penalized fit can overfit certain segments without providing a framework for selection

1. The peak may be over-fit, but it's hard to tell
2. The middle of the curve follows the noisy reversals
3. The tail jumps to chase noise – and the modeler is unsure of how much of the signal to recognize as real
4. Cross-validation indicates overfitting





Modeling a continuous variable with peaks, valleys, and tails

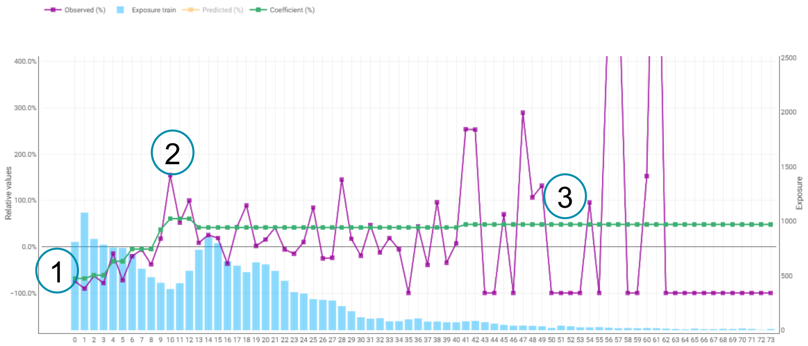
The optimally-penalized regression model systematically addresses the weak-points of the fit

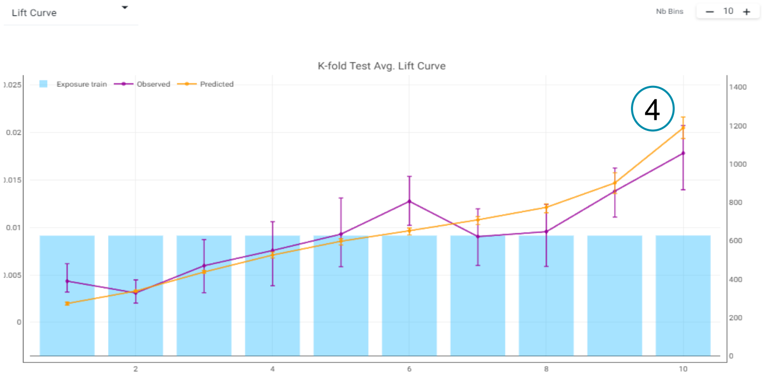
1. With high credibility in the younger homes, the fit nearly fully reacts to observed

2. Since the peak is at the exposure dip, the fit penalizes this point and does not over-react

3. The uptick in the older homes is recognized but to a small degree, due to the penalty – this gives the modeler a framework for recognizing some of the signal

4. Cross-validation shows better performance, less overfitting





Small Data - Categorical

Practical Scenario – Modeling on a small growing book

Scenario set up for mock homeowners’ book

The Set Up (Part 2):

Your product team ran a small pilot on a loss-saving water sensor device, and you’re asked to provide a data-driven *point estimate* for the benefit of this device

Note: all data displayed is a mock data set

Modeling a low-exposure categorical

Treatment for Categorical Variables in a GLM

Open question: How would you, the modeler, go about making a selection?

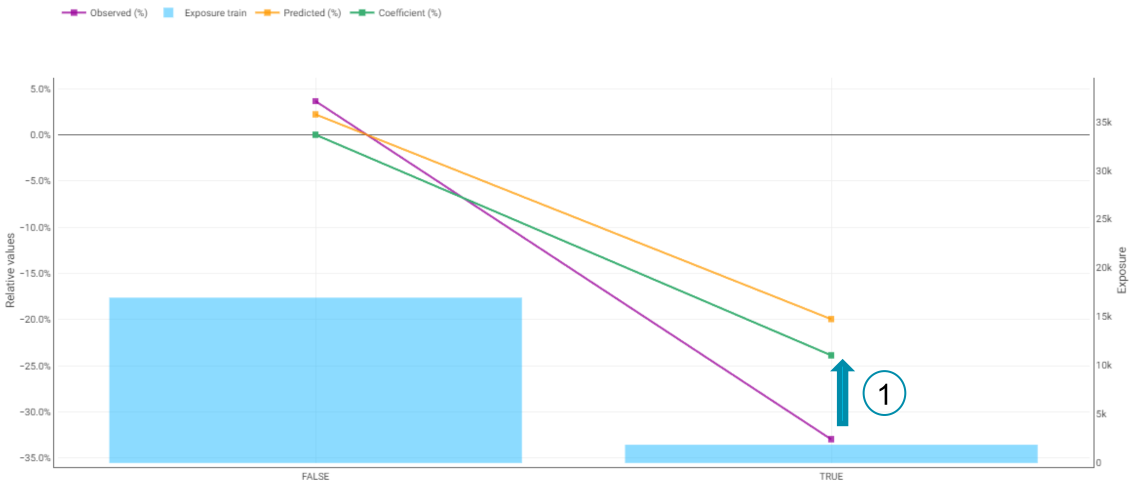
1. We only have a small amount of exposures with device installed
2. The signal is material and the GLM fully reacts to the signal
3. There is a wide confidence interval around the estimate

Water Sensor Installed (False, True)

Modeling a low-exposure categorical

Treatment for Categorical Variables in a penalized regression

1 The optimal penalized model gives us a credibility-weighed indication that automatically reduces the estimate



Recap: Tuning the penalty parameter to optimize performance

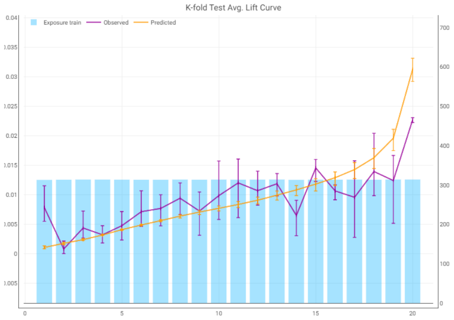
K-fold cross-validation illustrates how the penalty term can be tuned to optimize fit & performance

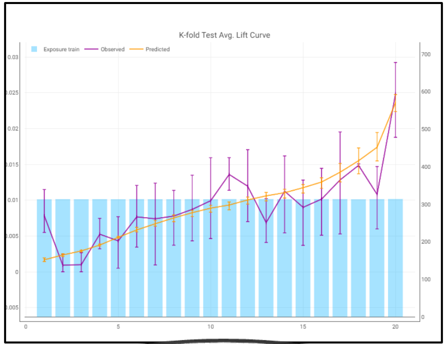
Least penalty | *Optimal penalty* | *Most penalty*

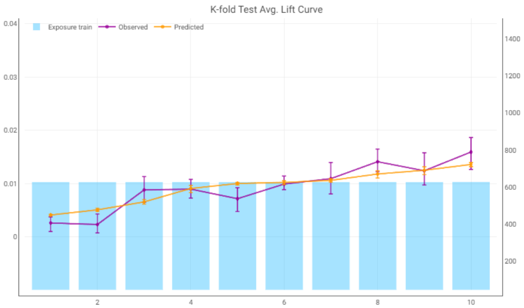
Too little penalty will overfit

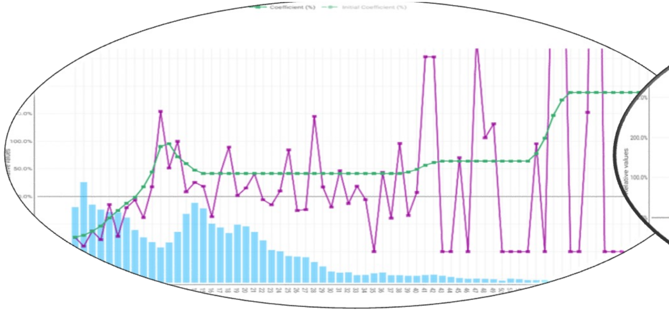
The higher penalty example is optimal

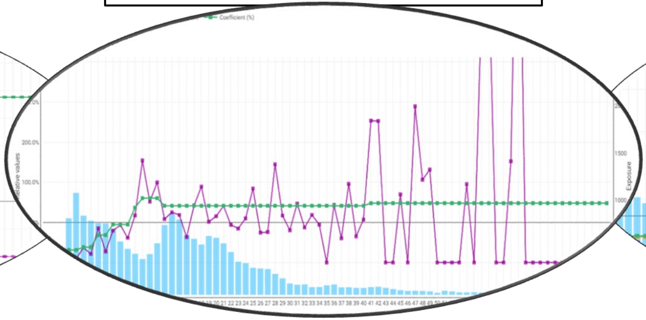
Too much penalty can result in smooth models that miss opportunity

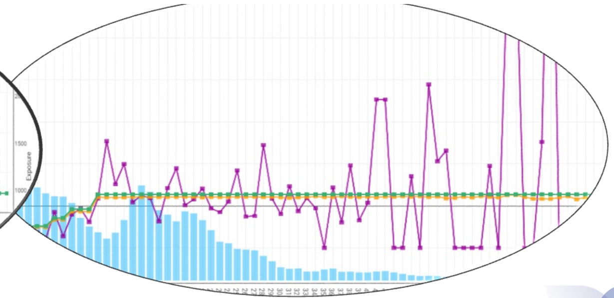












Small Data Scenario Recap – what did you achieve?

1. You’ve built a model that **recognizes the signal that matters**

2. Your model indications are credibility weighed consistently across variables, which sets you up to **step your pricing in the right direction** while allowing “room for growth” as more exposures come in

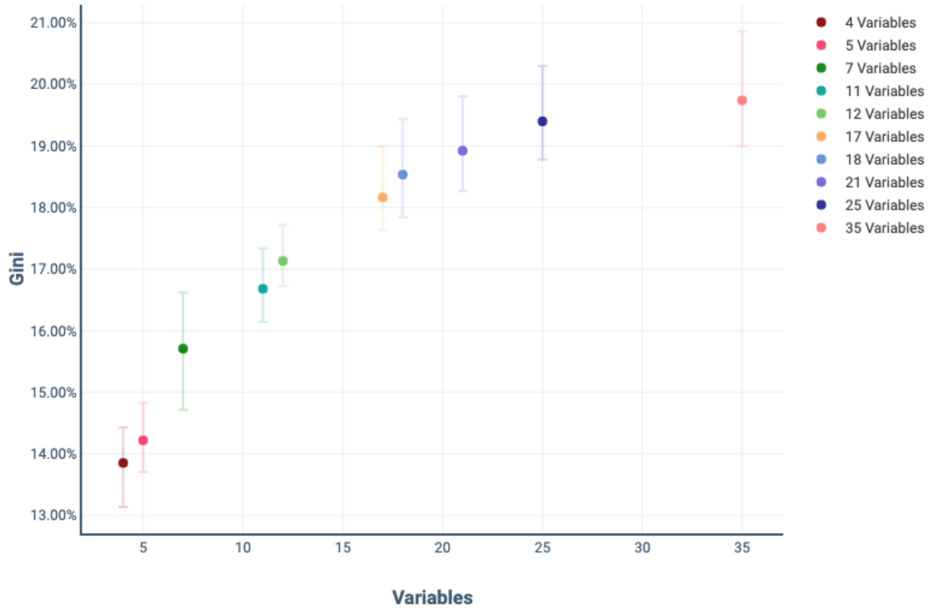
3. At an individual variable level, your model indications are responsive to the segment credibility, which helps you automatically draw the boundaries for **factor selections that are more stable** year over year and less likely to lead to future reversals

Variable Selection

Variable Selection Scenario

Practical Scenario

- **Practical scenario** – beginning stages of modeling
  - Dataset containing large number of predictors for initial consideration for use in model
  - Dataset contains multiple groups of correlated variables within pool of candidate variables
- **Goal** – reduce list of variables for consideration in the model
  - Tradeoff between total number of variables used and model performance – parsimony generally desired
  - Greater interpretability and ease of use with smaller variable list



Variable Selection Scenarios

Poll Question

How would you approach feature selection in this scenario?

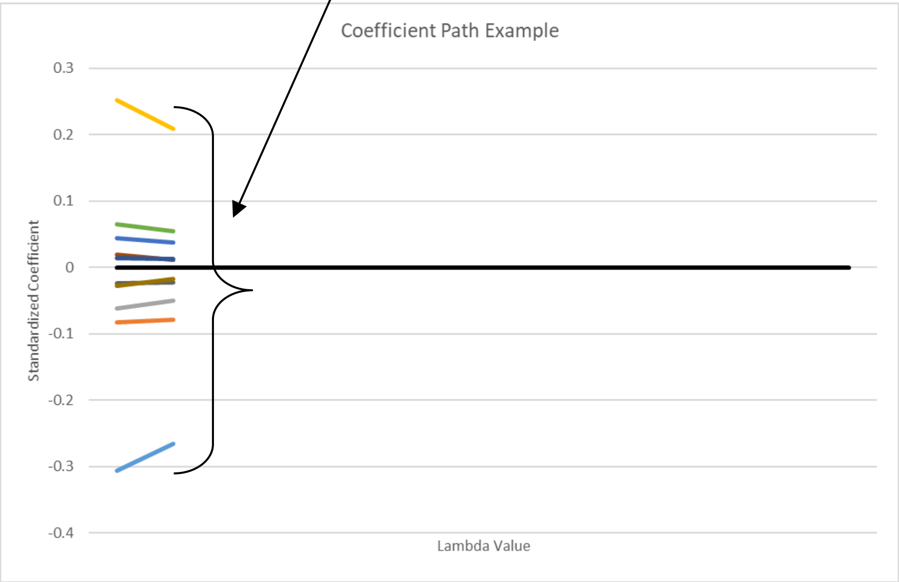
- Stepwise forward
- Start with variables “known” to be predictive, then testing new variables
- Throw it in a GBM and take the top N variables
- Penalized Regression (Lasso)
- Other?

Variable Selection Scenarios with Penalization

Selection using Lasso

- From example – looking at a pool of moderately correlated predictor variables
  - Look at variables in question across a range of lambda values
  - When lambda penalty zero/low, all/most variables will be included in the model with non-zero coefficient (**GLM type solution**)

Low penalty – all variables included

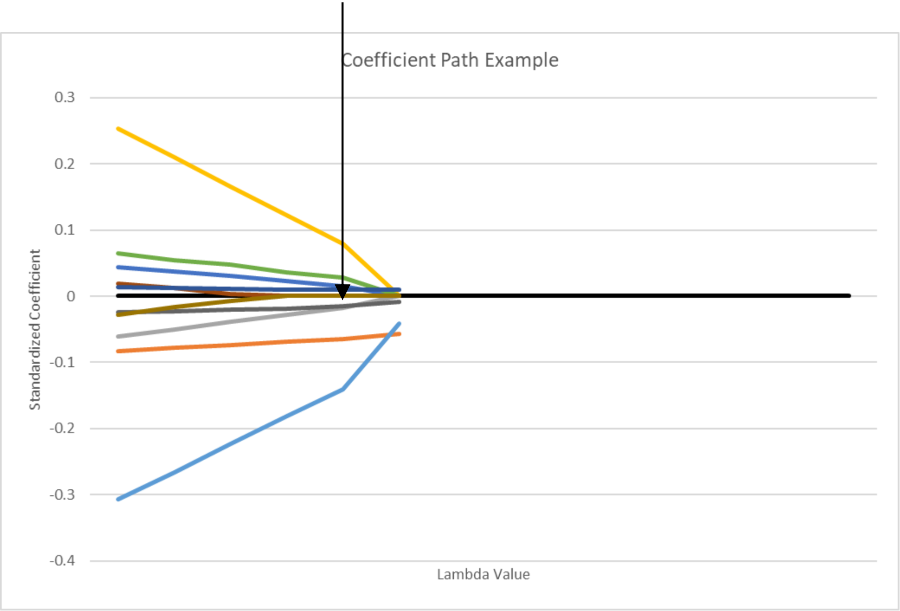


Variable Selection Scenarios with Penalization

Selection using Lasso

- From example – looking at a pool of moderately correlated predictor variables
    - Look at variables in question across a range of lambda values
    - As lambda penalty increases, many of the **correlated variables will quickly regularize out of the model**

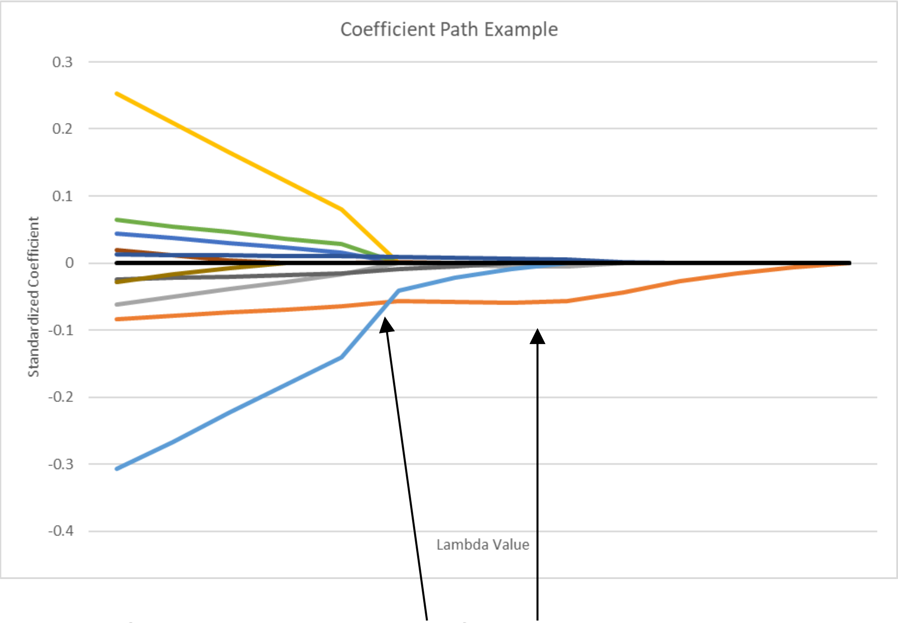
Many variables quickly reduce to 0



Variable Selection Scenarios with Penalization

Selection using Lasso

- From example – looking at a pool of moderately correlated predictor variables
    - Look at variables in question across a range of lambda values
    - Once lambda is high enough, only a few candidate variables will remain – **these are good candidates for closer examination in the model**

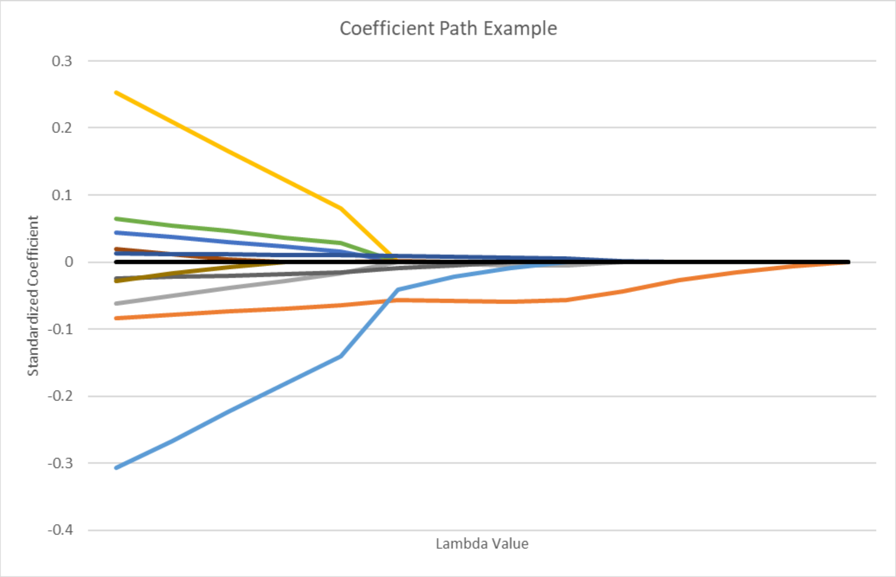


Candidate variables for selection – persist through higher values of lambda

Variable Selection Scenarios with Penalization

Selection using Lasso

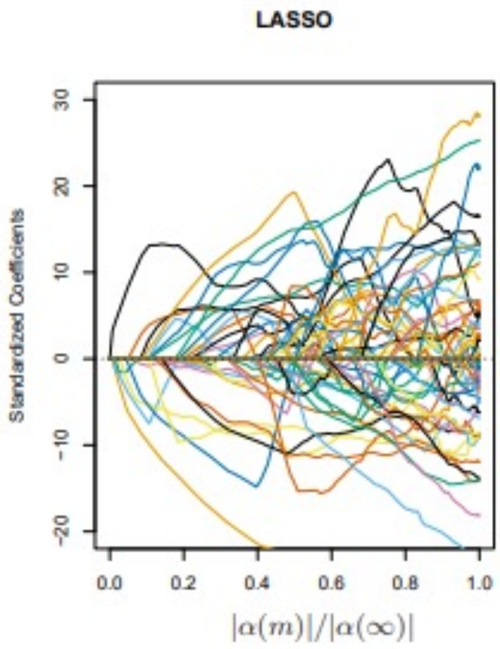
- With this method, variable selection using Lasso can:
  - Take advantage of ability of lasso to create **sparse solutions,** with many coefficients regularized to 0
  - Quickly select most important features from group of correlated variables
  - Extend to help in identification of **interactions** or best performing **transformations** (similar in principle to GAM framework)



Variable Selection Scenarios with Penalization

Selection using Lasso – Caution

- Very highly correlated predictors can present some challenges
  - Lasso can be somewhat indifferent among groups of nearly collinear variables
  - This can cause variables to come in/out of model across range of lambdas, and **cause coefficient paths to not be strictly monotonic**
  - Graphic shown is an extreme example, but in practice you may not always see clean, monotonic variable progression
  - **Solution** – run correlation matrix, understand groups of variables with extremely high correlation



https://hastie.su.domains/Papers/ESLII.pdf, 16.2.3

Near Aliasing

Model Stability

GLMs and Correlation

Lower Correlation ↔ *Higher Correlation*

Low to moderate correlation
Strength of GLMs in being able to separate out unique effect of each variable

✓

High Correlation (near aliasing)
Not always identified by modeling software, can result in erratic coefficients and cause issues with model stability

Perfect Correlation (Aliasing)
To be avoided, but removed automatically by most modeling software

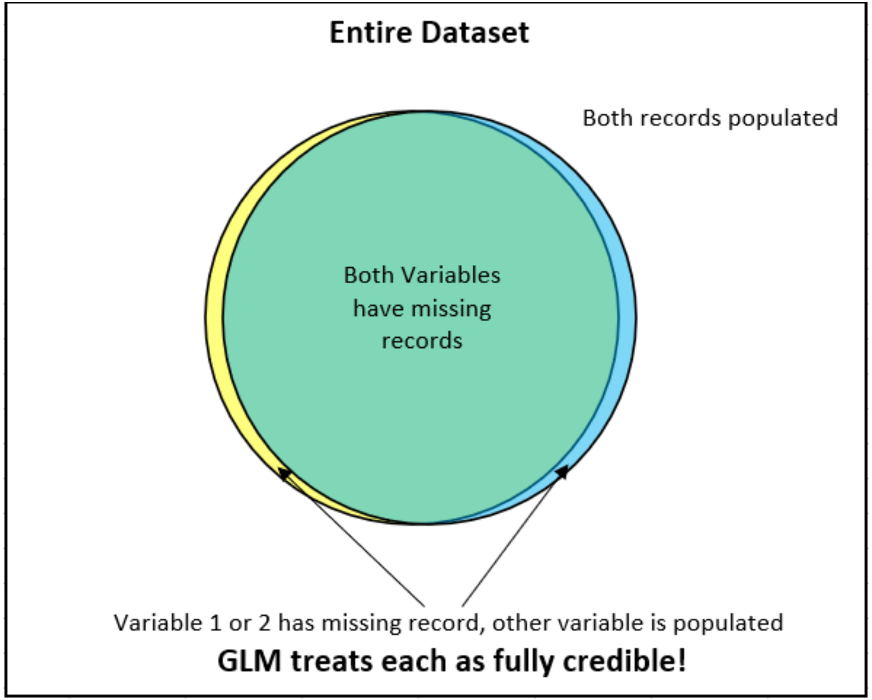
Nearly-Aliased Variables without Penalization

Practical Example

- We want to generate relativities for 2 different variables
    - Each variable is drawn from the same data source
    - Each variable also has missing records – error flags also included to control for missing information
    - Variables have moderate correlation, but **error flags are highly correlated**

      (>0.99 correlation coefficient)

How would you handle the near aliased variables?

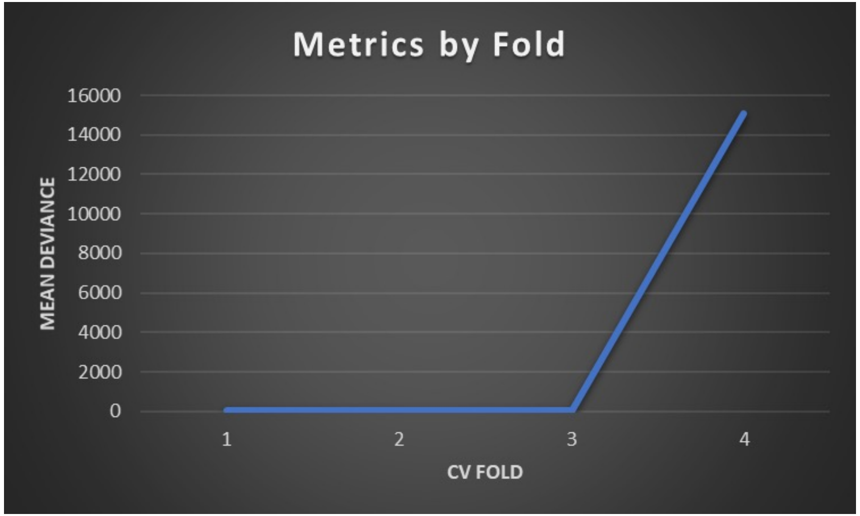


Nearly-Aliased Variables without Penalization

Model stability Issues

- Simple GLM model fit with 4-fold Cross-Validation to generate relativities
  - Model fit produced with just the 2 variables in question and their associated error flags
  - 2 additional control variables also included in the model
  - Result – model instability and loss of interpretability on metrics, univariates

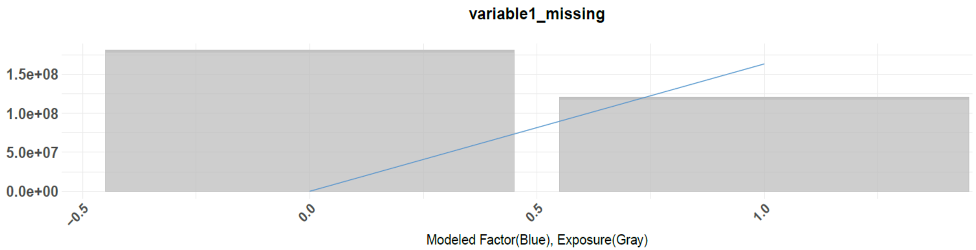
- **What happened in fold 4?**

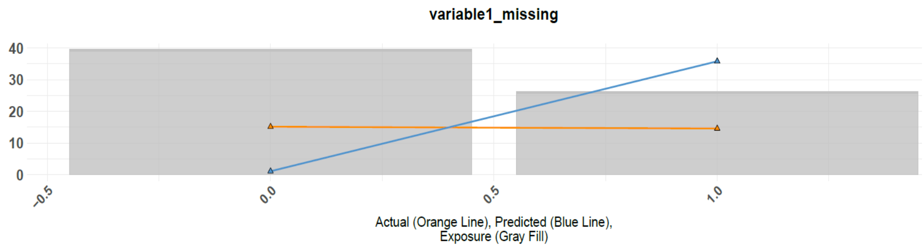


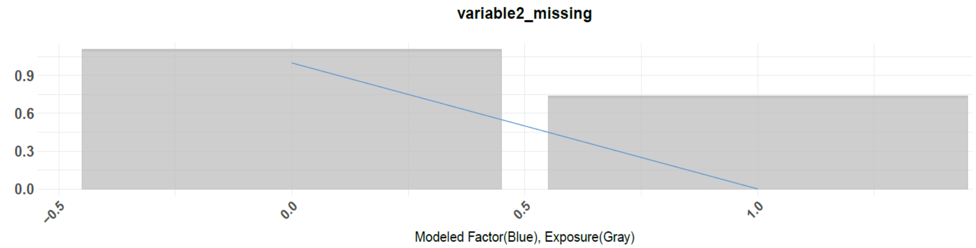
Nearly-Aliased Variables without Penalization

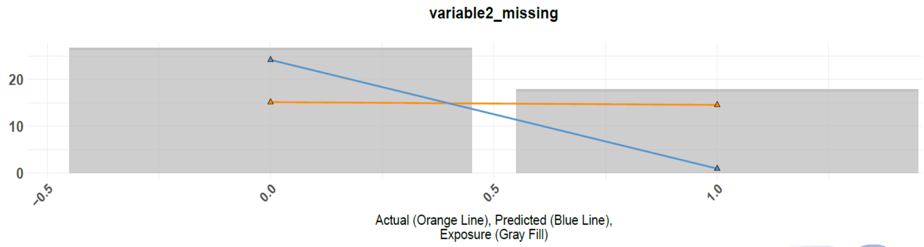
Model stability issues – Closer look at fold 4

- Nearly aliased variables driving instability in error flag variables - Large coefficients in opposite direction, with high standard errors
- Small shift in data across folds causing extreme coefficient swings and lack of interpretable output





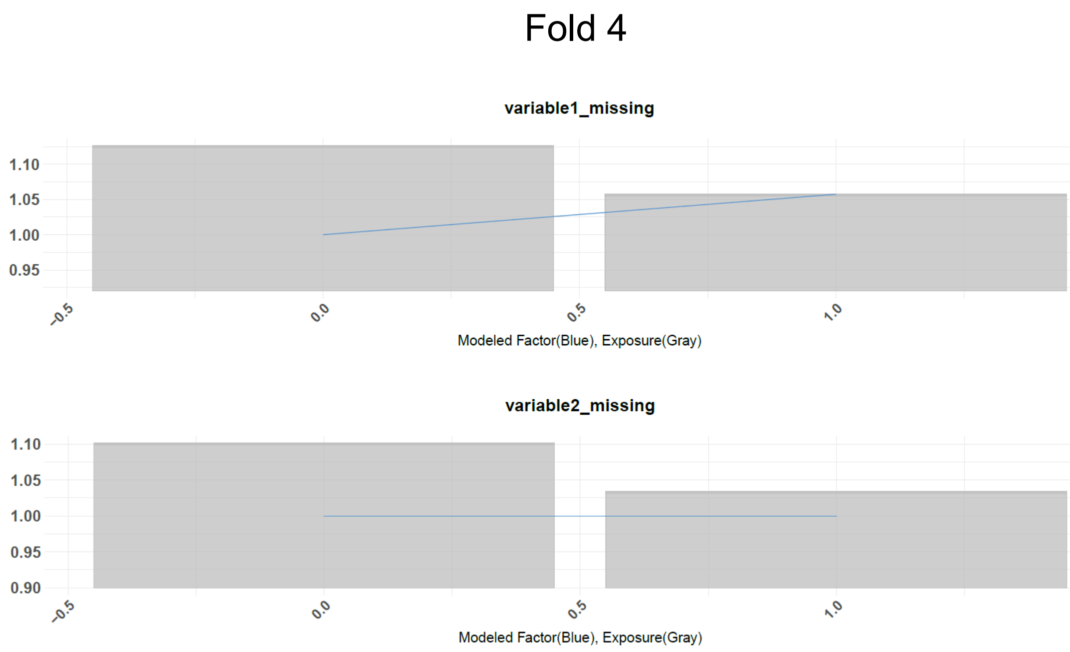




Nearly-Aliased Variables with Penalization

No more stability issues

- Small penalty term introduced to the model
  - One missing flag regularizes out
  - Avoids trying to assign unique signal to each missing flag

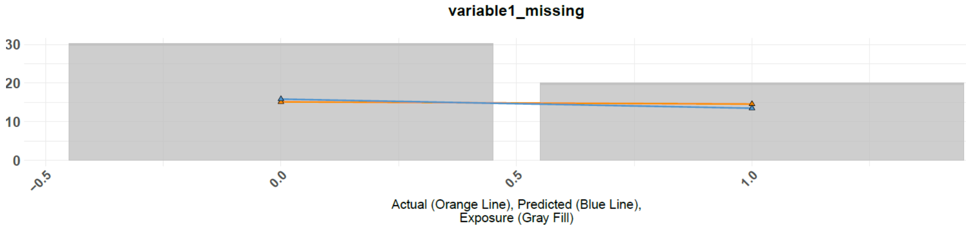


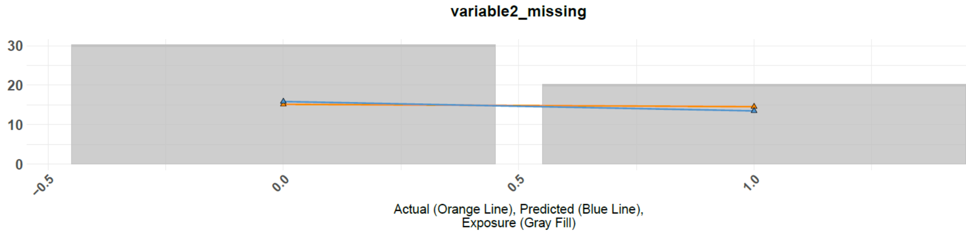
Nearly-Aliased Variables with Penalization

No more stability issues

- Multiple methods can be used to address high correlation
  - Regularization can automatically decide which level to drop instead of manual decision
  - Regularization is often preferred method of addressing collinearity issues
  - Example can be extended to any near-aliased modeled predictors
- Not a replacement for understanding correlations of variables prior to modeling!
  - But not always easy to catch issues arising from nearly aliased variables or multi-collinearity up front

Fold 4





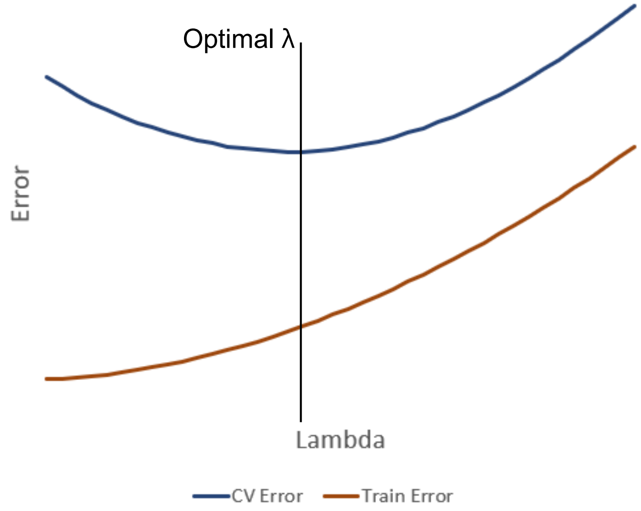
Conclusion

Modeler Resources

Additional Cost for Penalization?

- What about the additional cost of the model build when using penalization?
    - Two additional tuning parameters introduced, alpha and lambda
    - Even if we assume Lasso form (i.e. fix alpha at 1), still additional lambda parameter to tune
    - Lambda typically optimized using Cross-Validation, adding compute time
- **Yes, but additional cost is minimal**, and efficiencies can be gained when adding regularization even when using large, ‘highly credible’ dataset

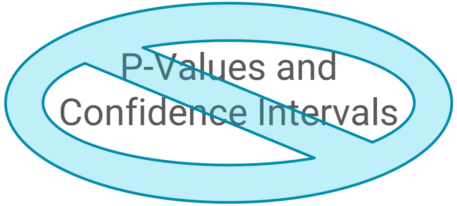
$$\lambda\left(\underbrace{\frac{1-\alpha}{2}\sum_{j=1}^{m}\hat{\beta}_j^2}_{L_2\text{ (Ridge)}} + \alpha\underbrace{\sum_{j=1}^{m}|\hat{\beta}_j|}_{L_1\text{ (Lasso)}}\right)$$

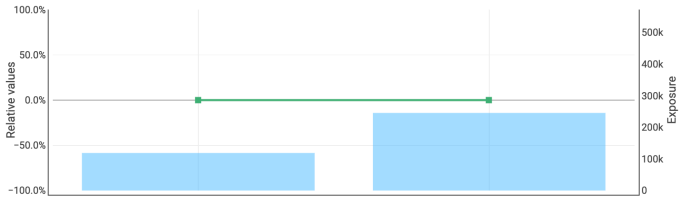


Modeler Resources

Additional Cost for Penalization?

- **Variable selection is automated**, reducing time analyzing borderline variables that are subject to actuarial judgment and groupings





- Automatic variable removal promotes **consistency in variable treatment across different modelers given similar modeling situations,** speeding up time in standardizing decisions across a modeling team

Variable Treatment From Modeler1 ⟷ Variable Treatment From Modeler2 ⟷ Variable Treatment From Modeler3

- **Flexibility to fit standard GLM** if lambda value of 0 is optimal result in tuning process

Penalized Regression has Practical Impact