Comparison of Penalized Modeling Techniques

NAIC Predictive Modeling Book Club - April 2025

AKUR8

Thomas Holmes

Chief Actuary for the US Region

Agenda

GLM + Credibility

Lasso and Ridge

Musical Interlude

MGCV GAM and Derivative Lasso

Lasso Credibility

First, Non-Penalized GLM

AKUR8

GLMs and Low-Exposure Levels

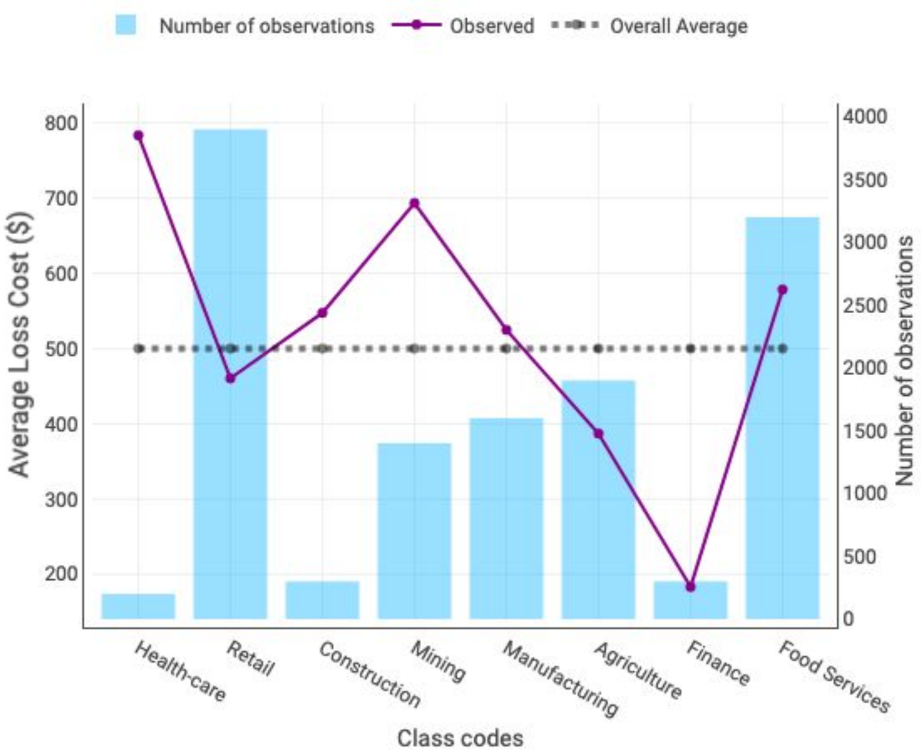
Worker's Compensation example

Blue bars = Exposure

Purple lines = Observed Loss

Black line = Overall Average Loss

Observed loss by class code



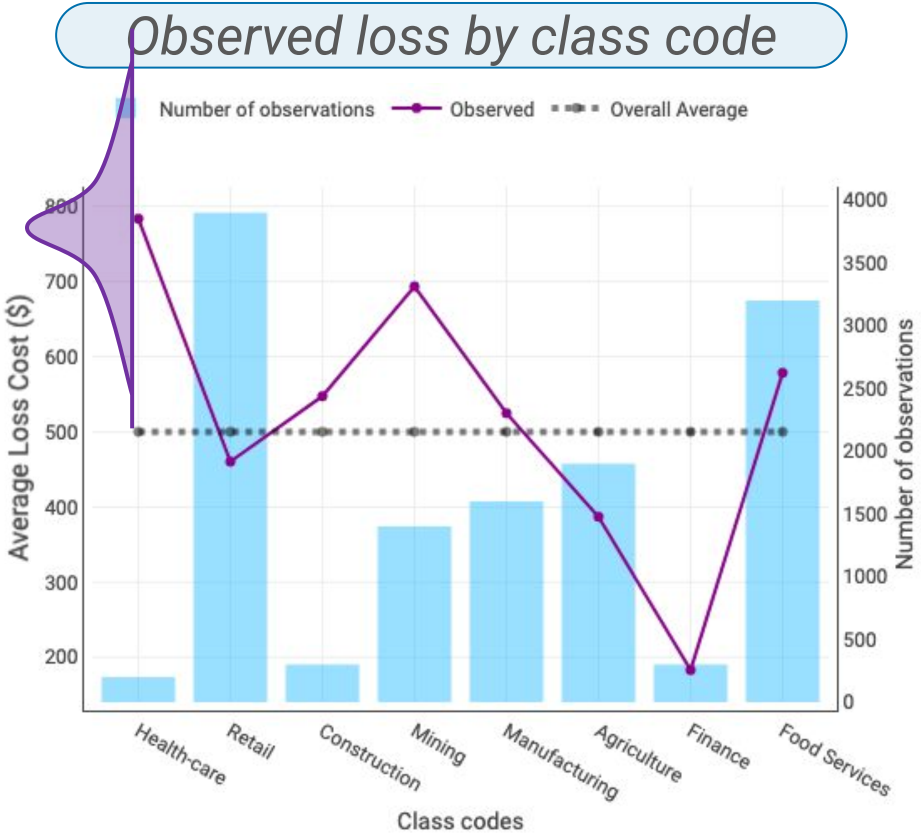
AKUR8

Worker's Compensation example

GLMs give full credibility to the observed data.

In this example, GLM statistics reflect the uncertainty of the observed relativities as the "real" relativities.

Bell curve represents the 90% confidence interval.



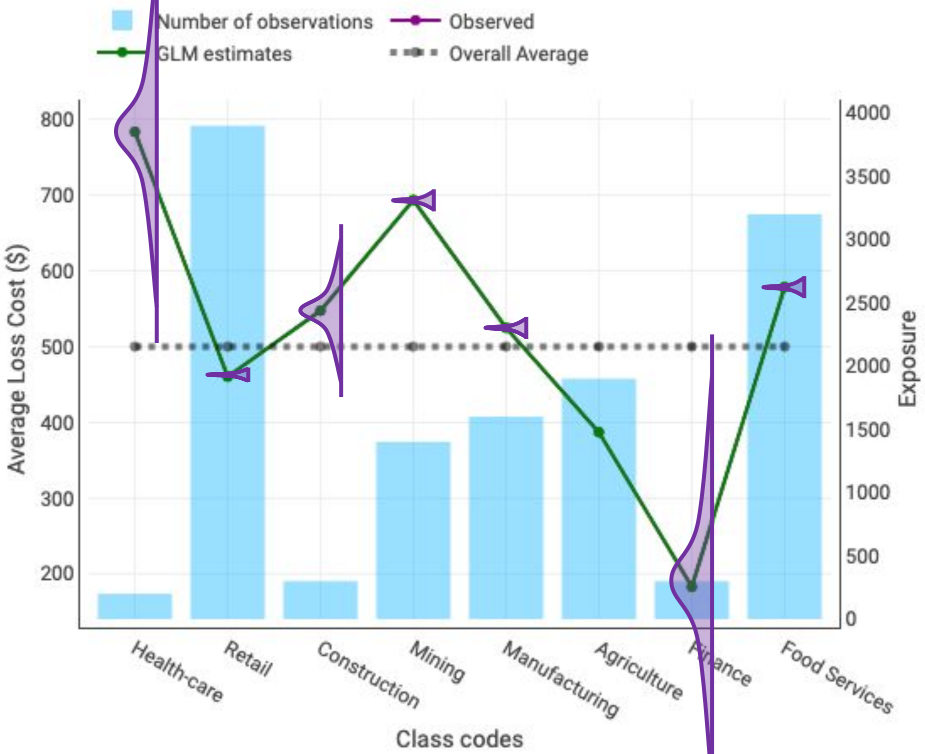
AKUR8

Worker's Compensation Example

The purple bell curve represents the confidence interval for the significance of each coefficient.

A common significance threshold is .05 - this is P-value review.

Estimate of loss by class code



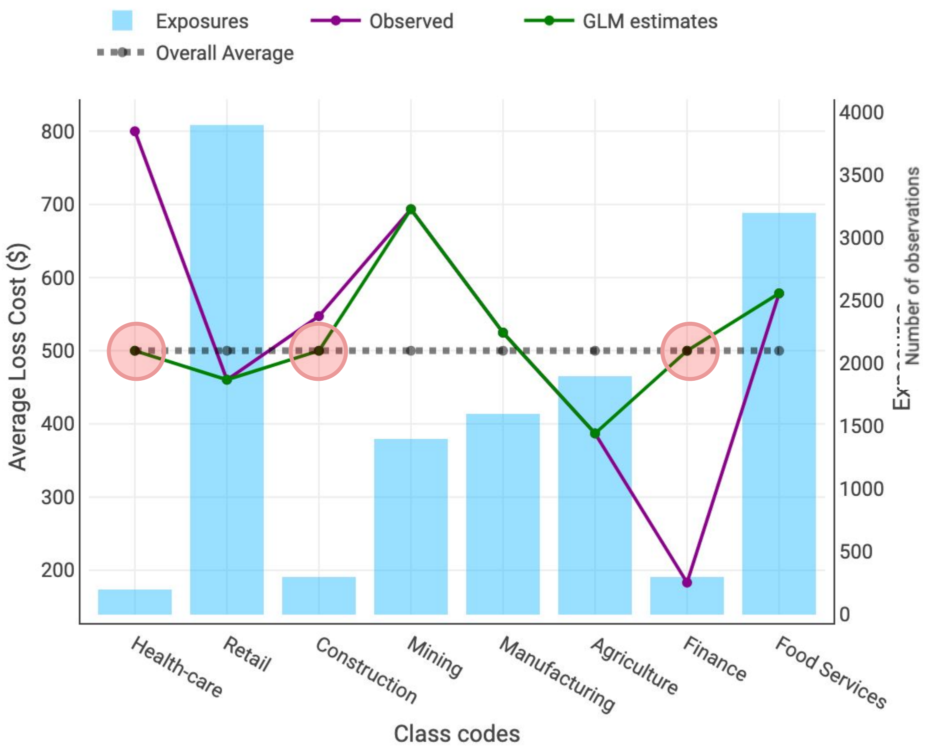
AKUR8

Worker’s Compensation Example

We remove the insignificant coefficients and create a new model.

These new estimates completely remove the differentiation for these categories.

Estimate of loss by class code

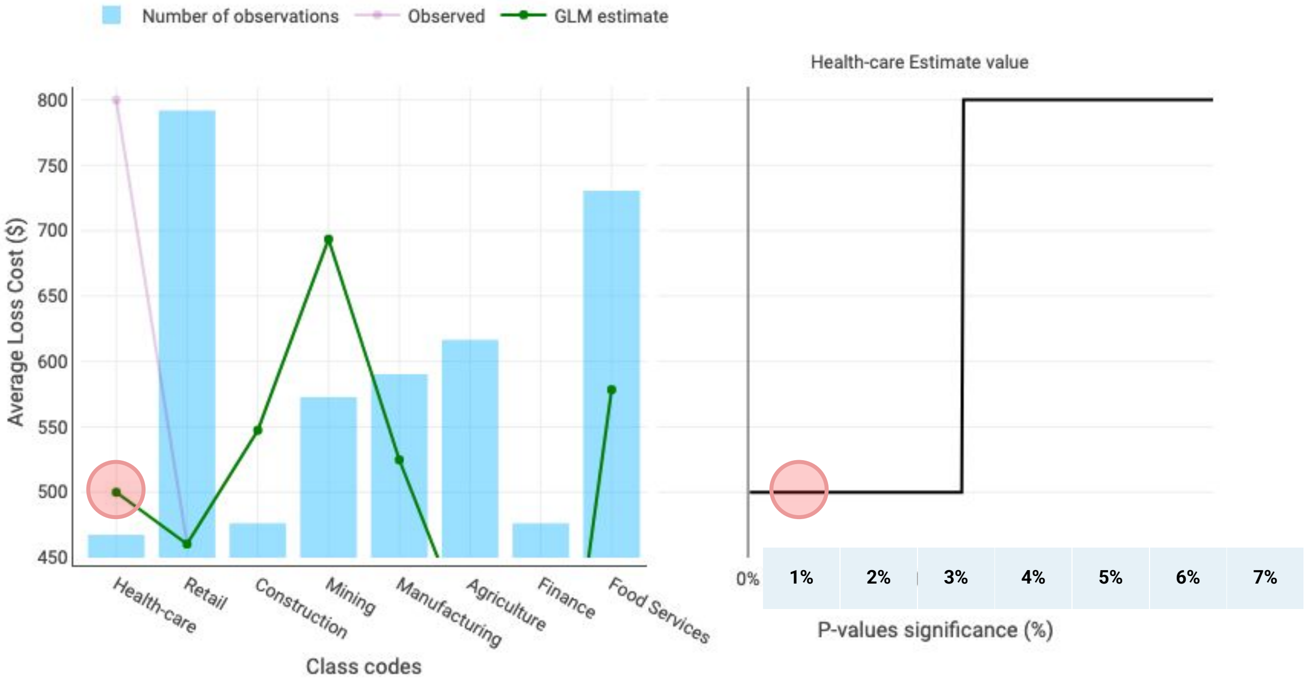


AKUR8

The Significance Threshold is Very Important

Strong significance thresholds lead to a **robust** model.

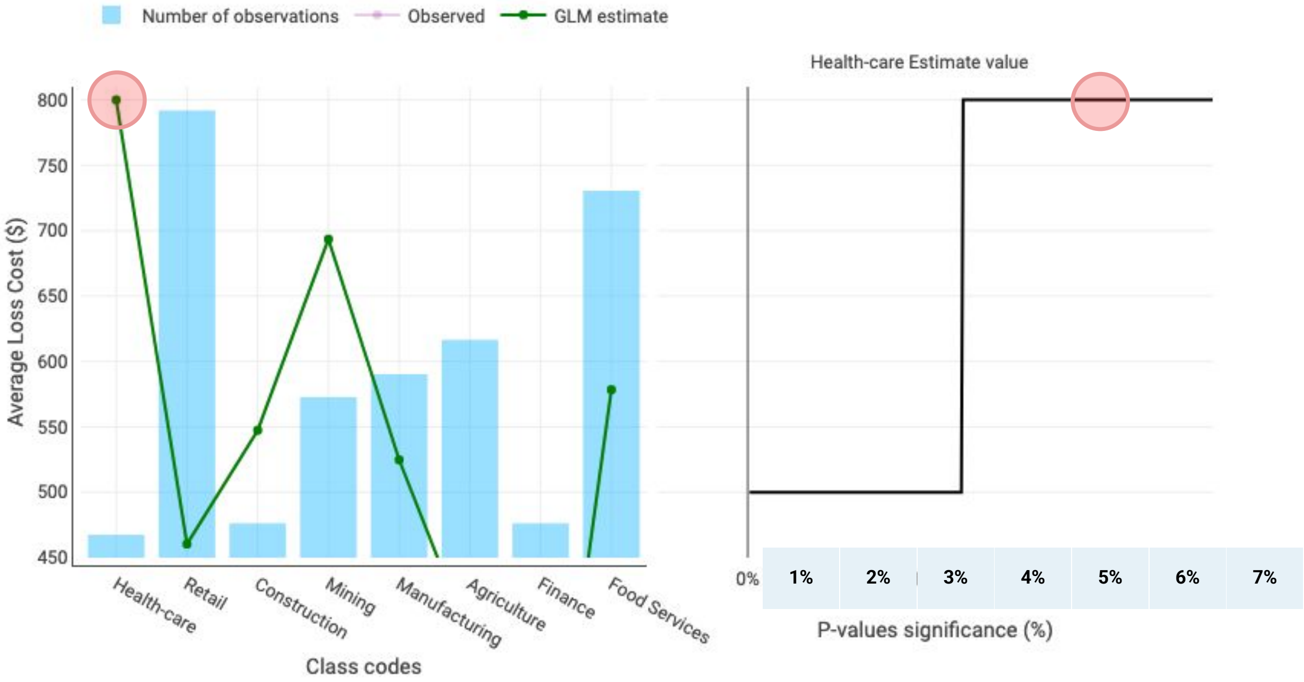
Small changes to the data are unlikely to cause a material shift in a robust model.



The Significance Threshold is Very Important

Weak significance thresholds lead to a **volatile** model.

Small changes to the data may cause a material shift in a volatile model.



GLM
vs
Penalized Regression

AKUR8

Significance
vs
Credibility

AKUR8

Interpreting penalization as credibility can aid in the actuarial review of penalized models

AKUR8

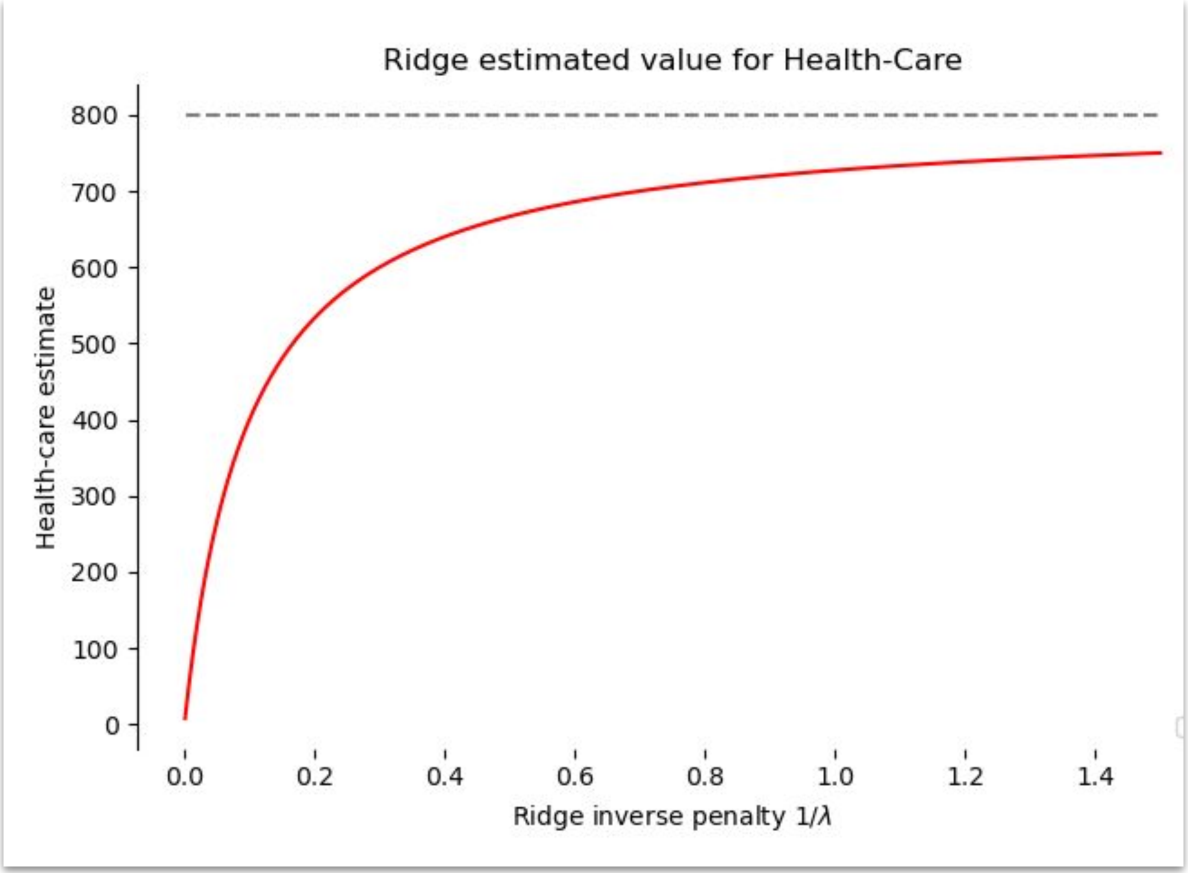
Not all penalization is actuarial credibility, but the perspective is extremely helpful.

AKUR8

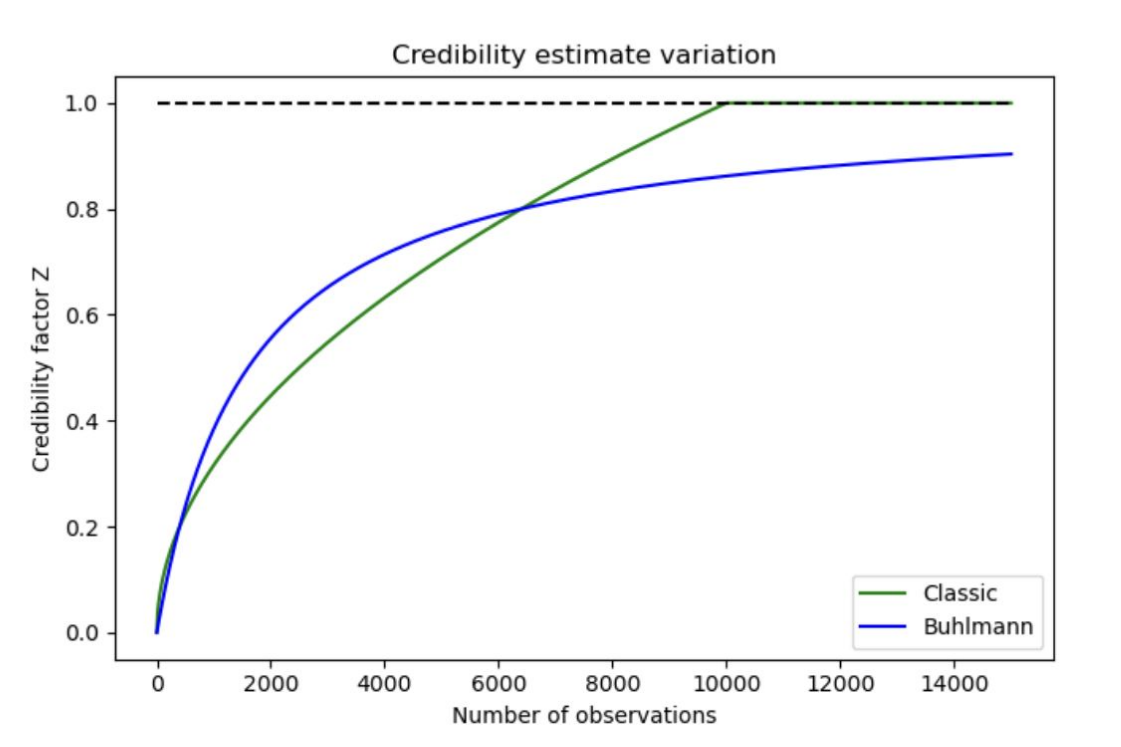
Let’s start with Ridge

AKUR8

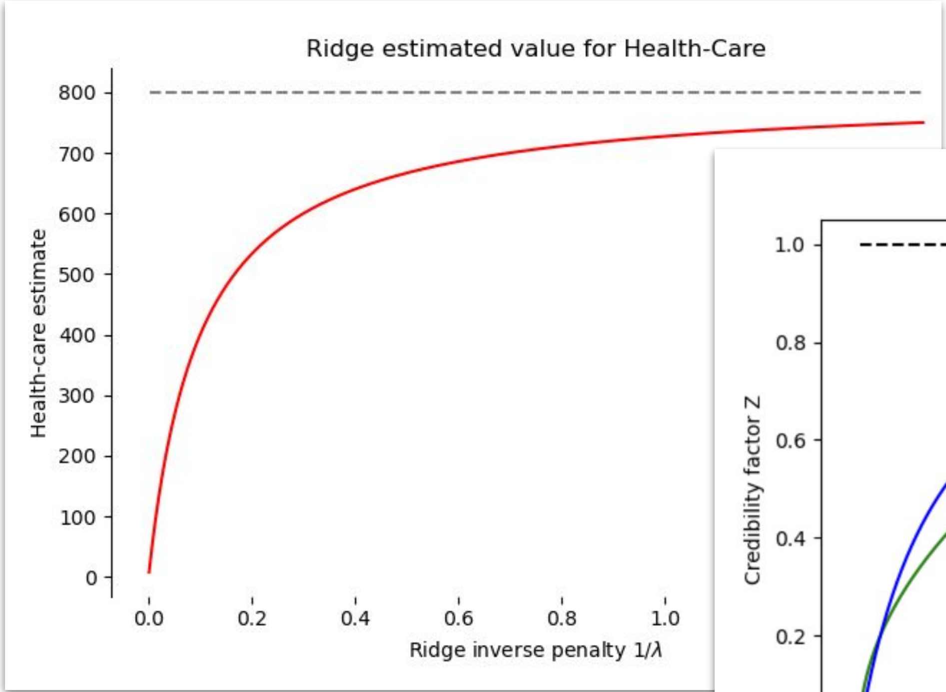
Ridge Penalization - The Effect of the Penalty Lambda (λ)

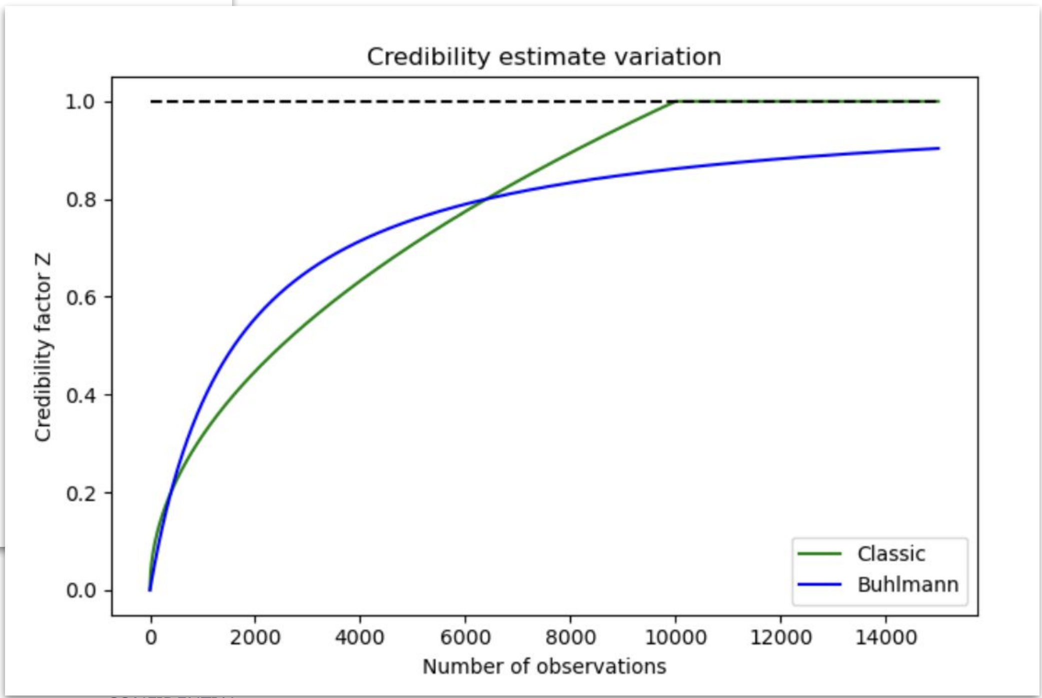


Buhlmann and Classical Credibility

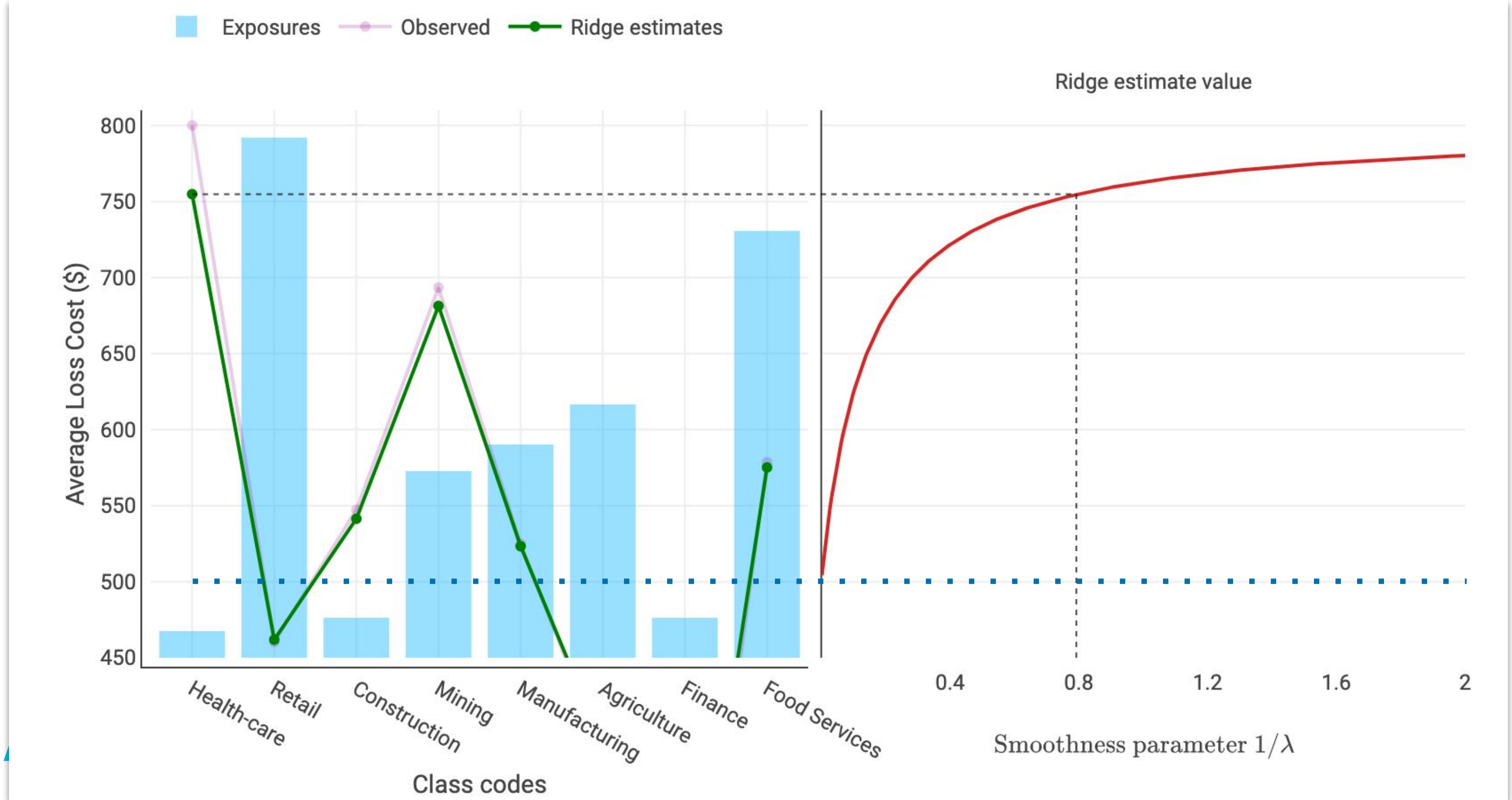


These seem similar ....

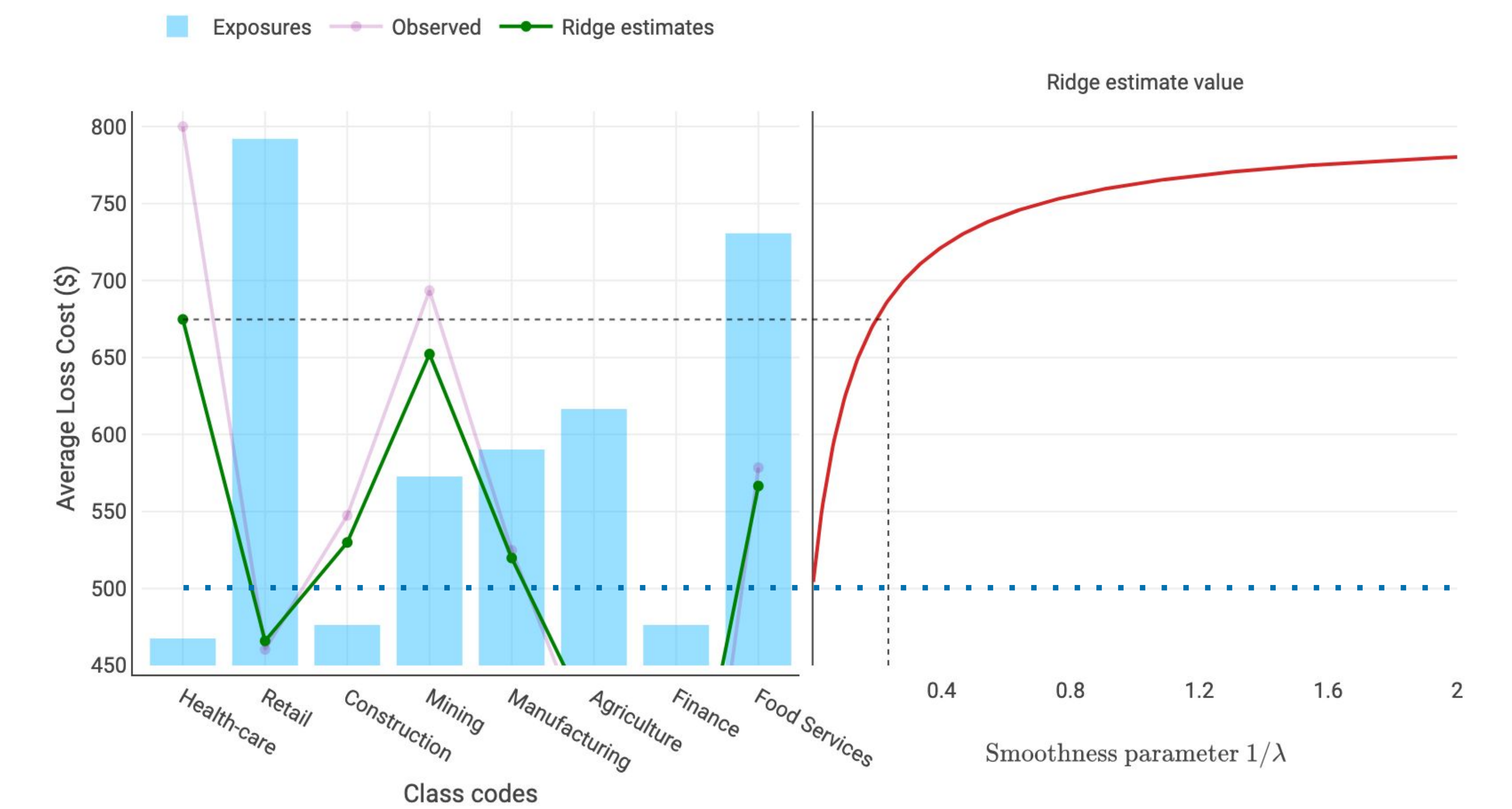




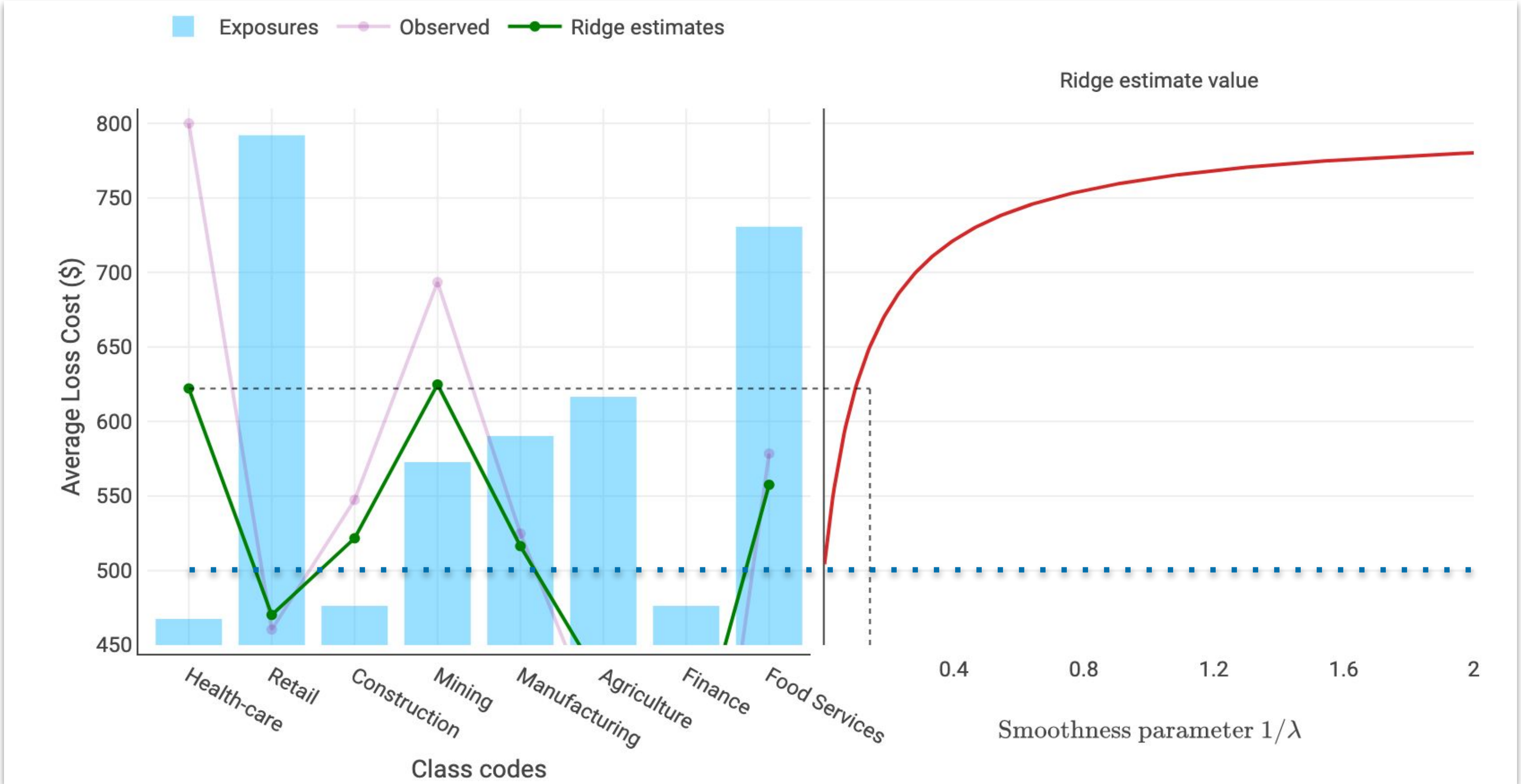
Ridge Health Care Estimate: Small λ



Ridge Health Care Estimate: Medium λ



Ridge Health Care Estimate: Large λ



Coefficient Path for Ridge Regression

Credibility vs. Significance

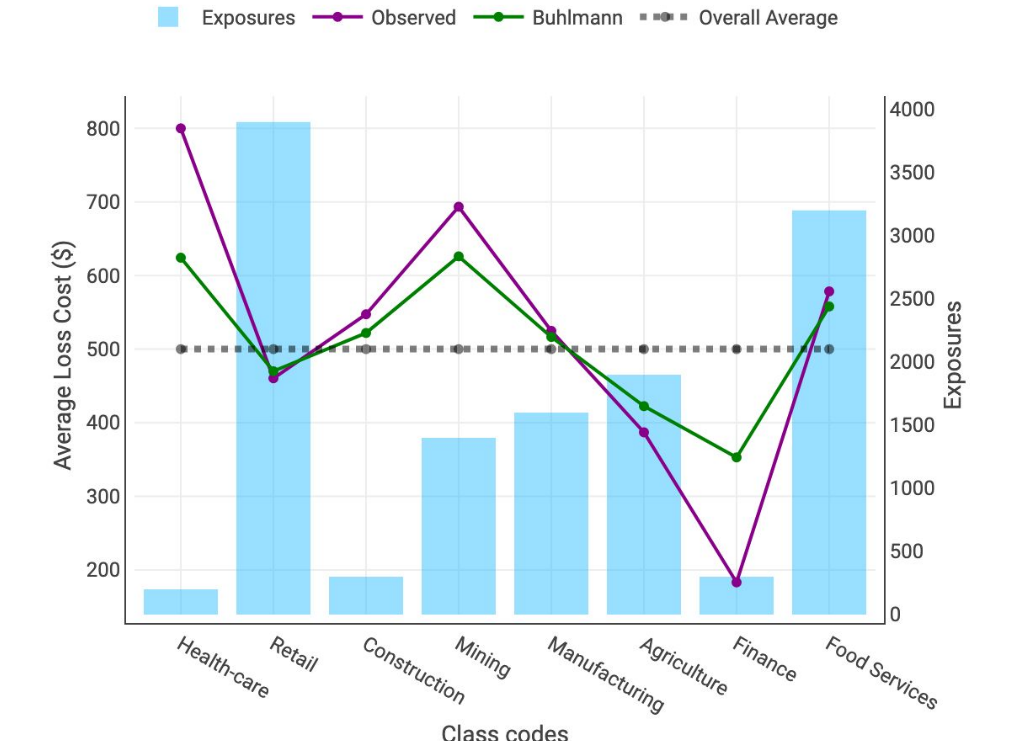
Low exposure levels are:

Not fully trusted

Not fully discarded

AKUR8

Estimate of loss by class code



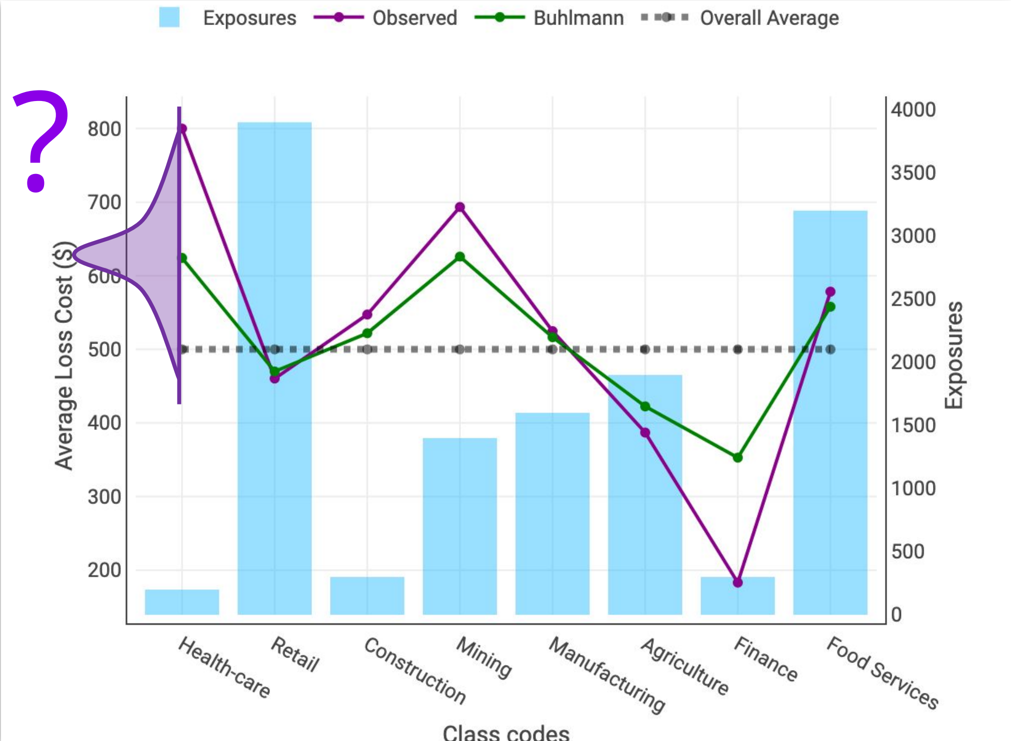
Credibility vs. Significance

Low exposure levels are:

Not fully trusted

Not fully discarded

Estimate of loss by class code



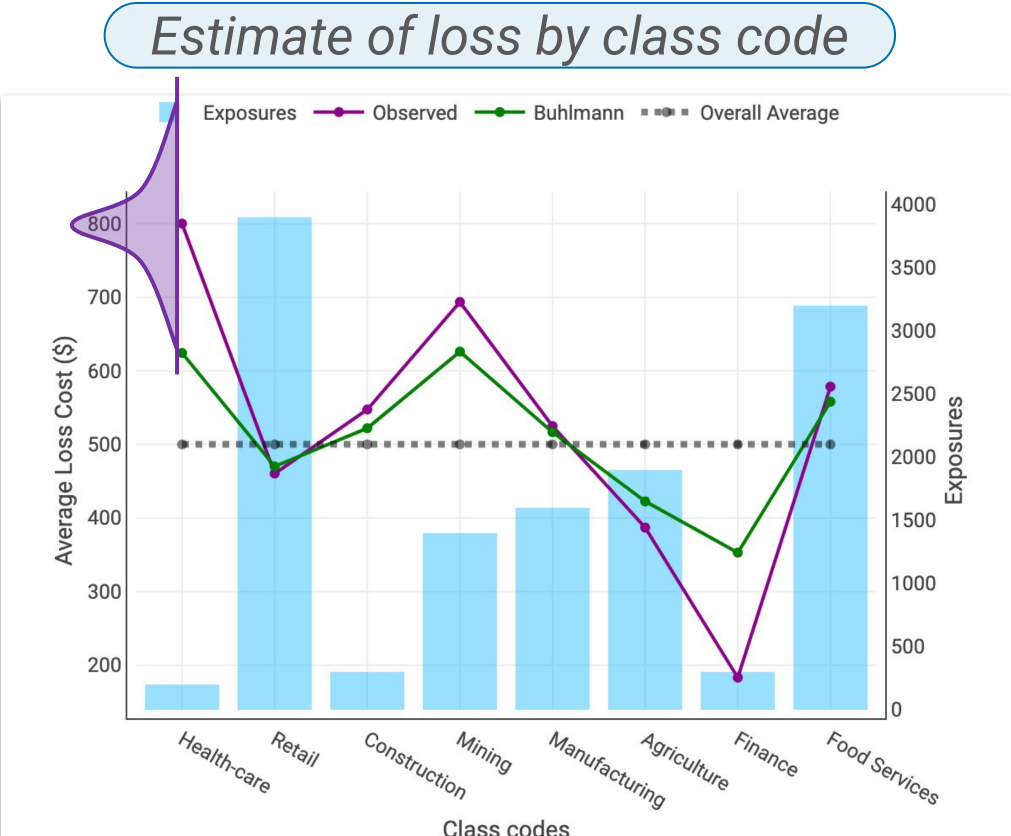
AKUR8

Credibility vs. Significance

Penalization weighs between the observed and overall average to estimate the “most likely” set of coefficients.

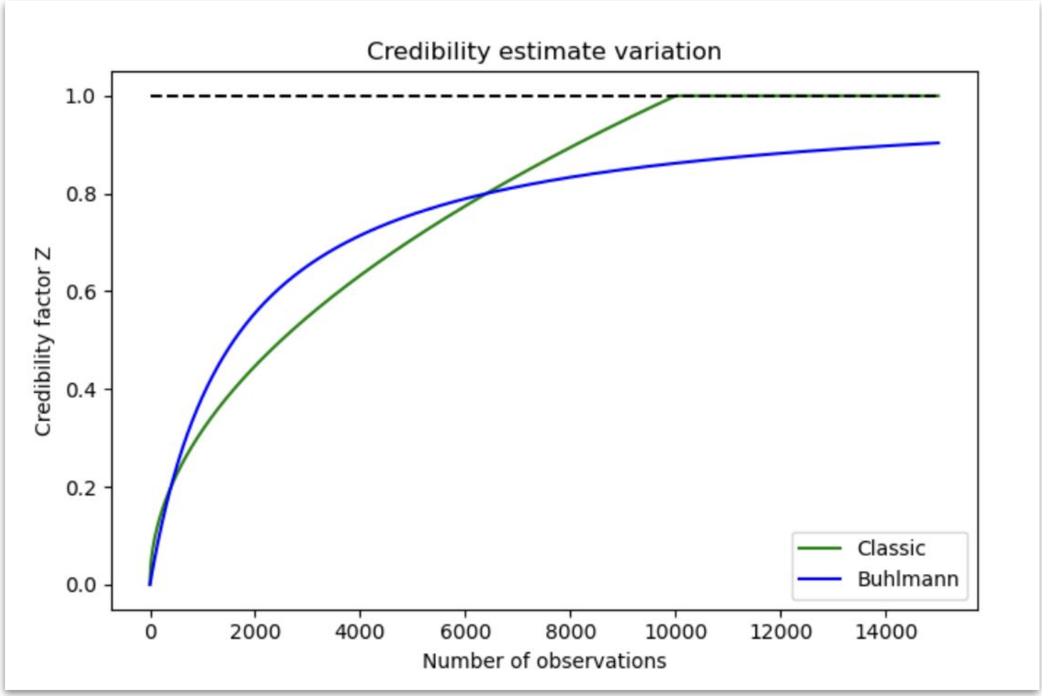
Different types of penalization perform this weighting differently.

Credibility-like shrinkage is **likelihood-based.**



AKUR8

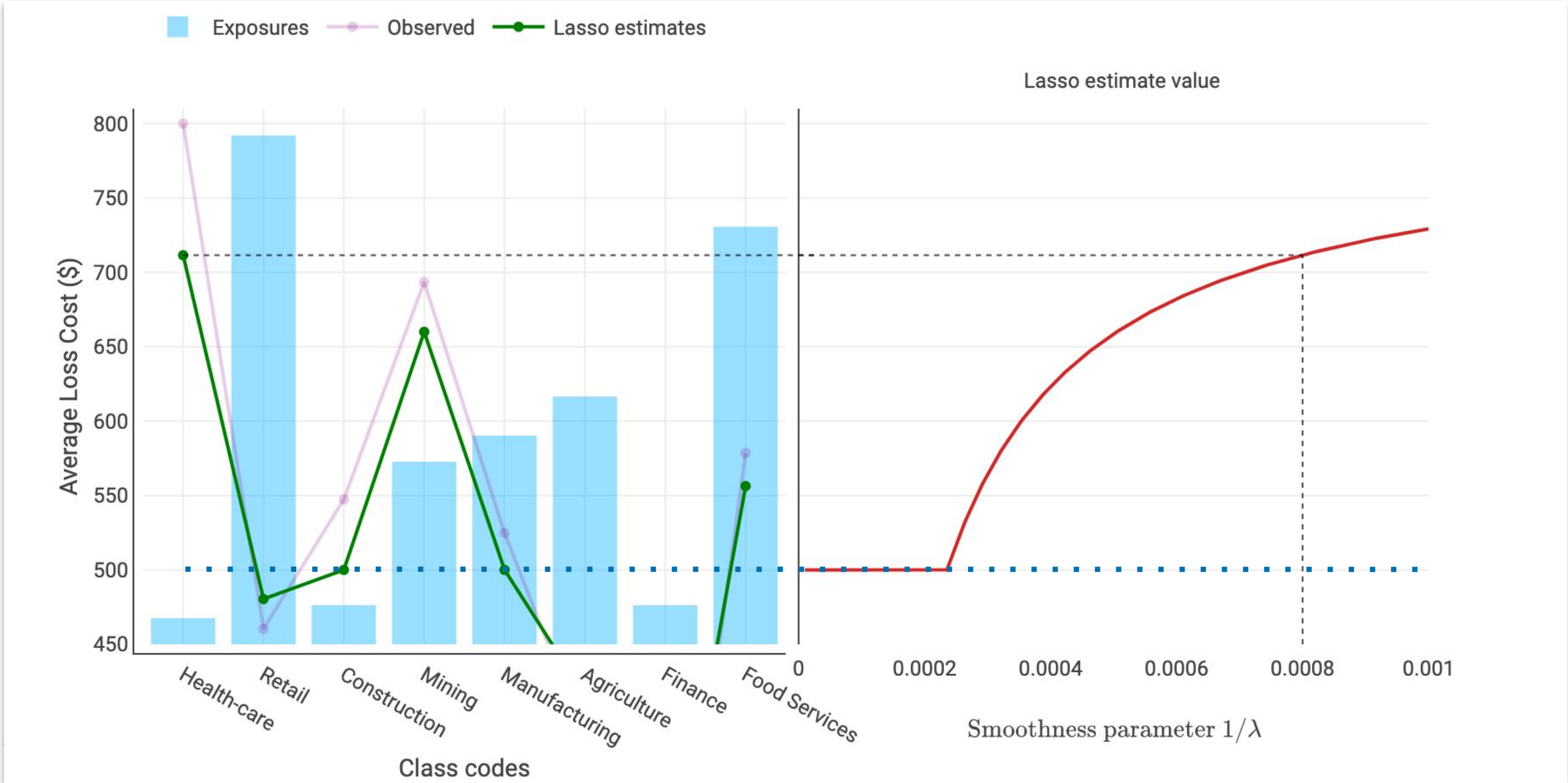
Classical and Buhlmann also Apply Credibility Differently



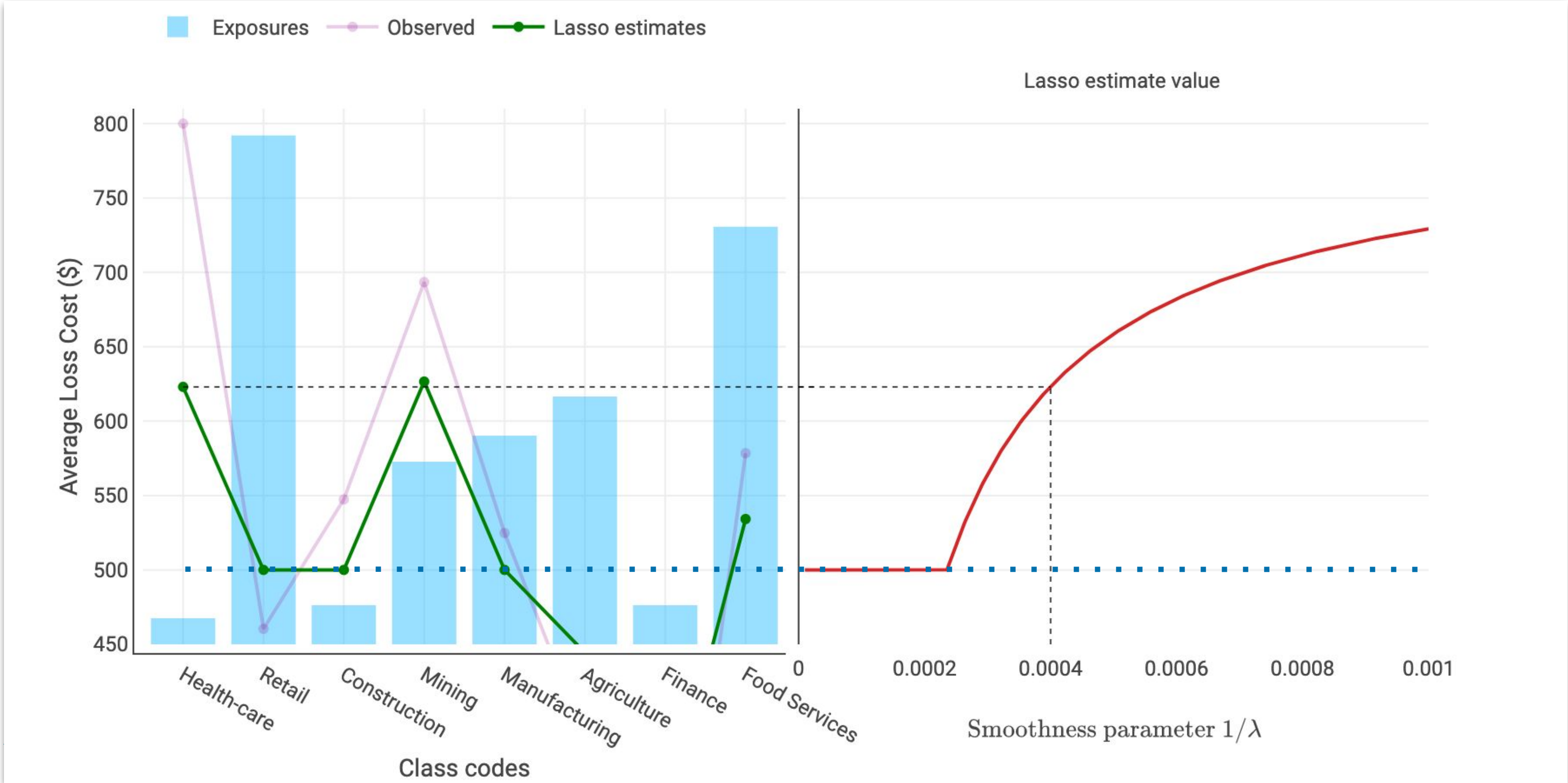
Lasso

AKUR8

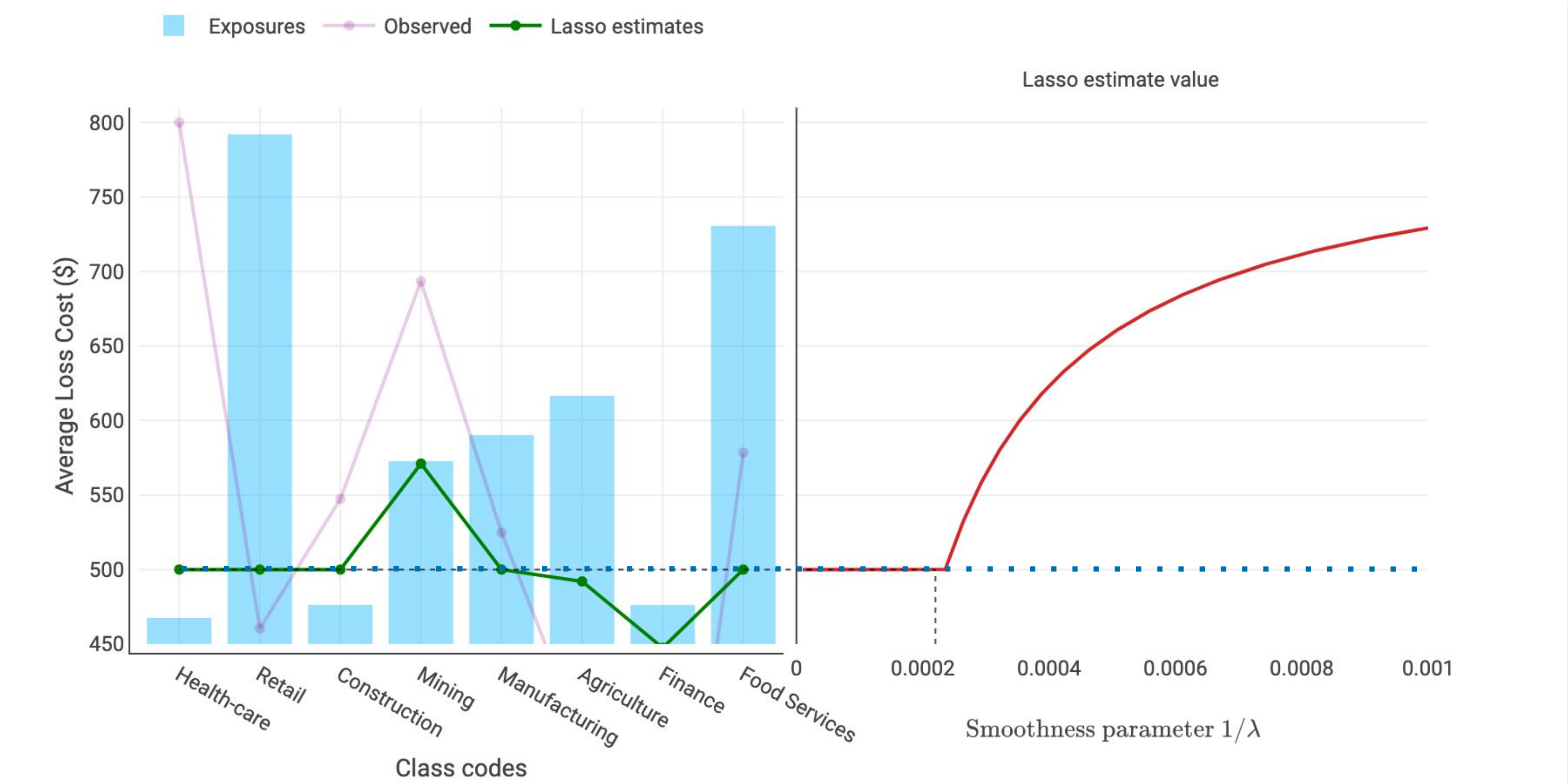
Lasso Health Care Estimate: Large λ



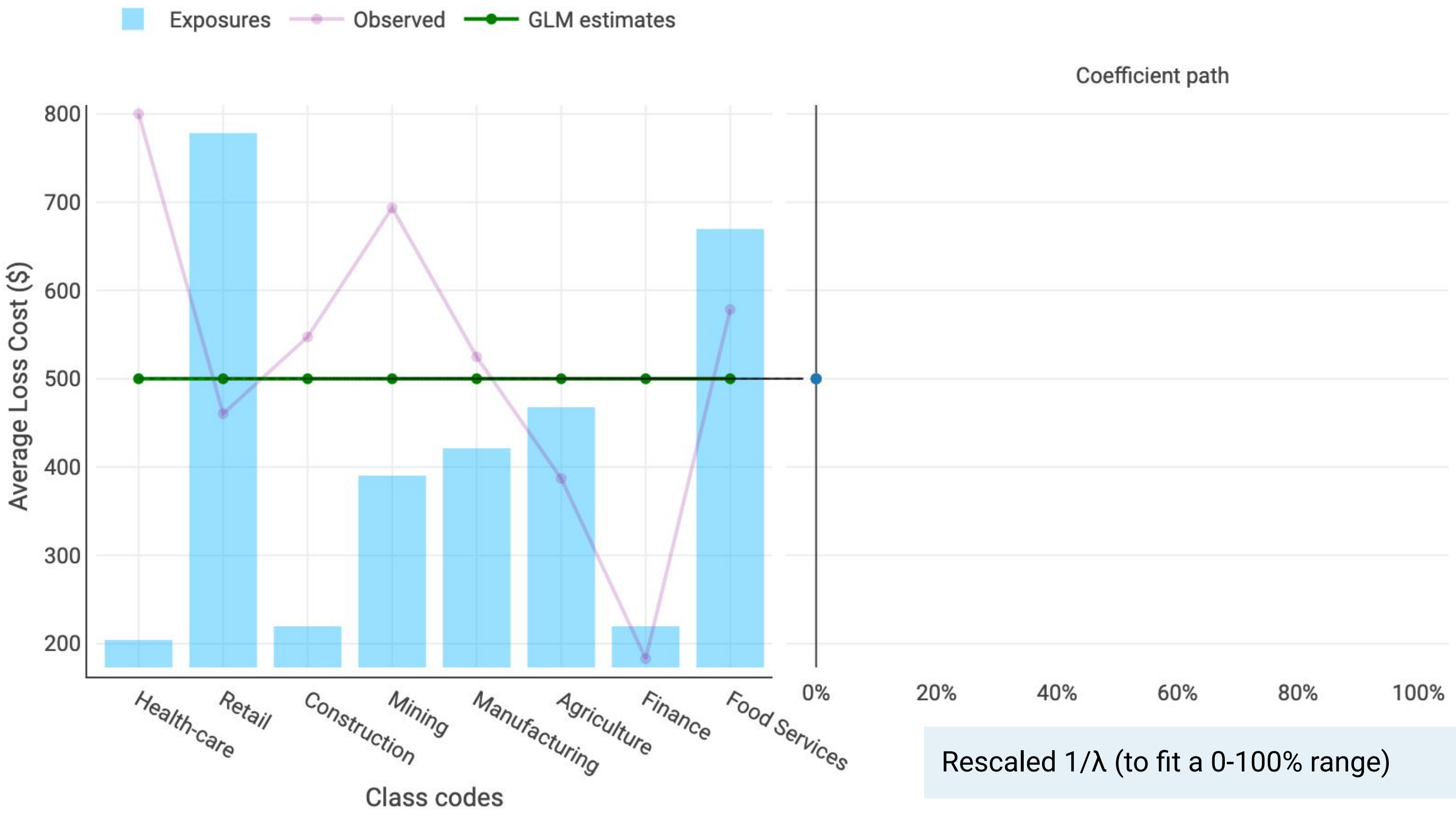
Lasso Health Care Estimate: Medium λ



Lasso Health Care Estimate: Small λ

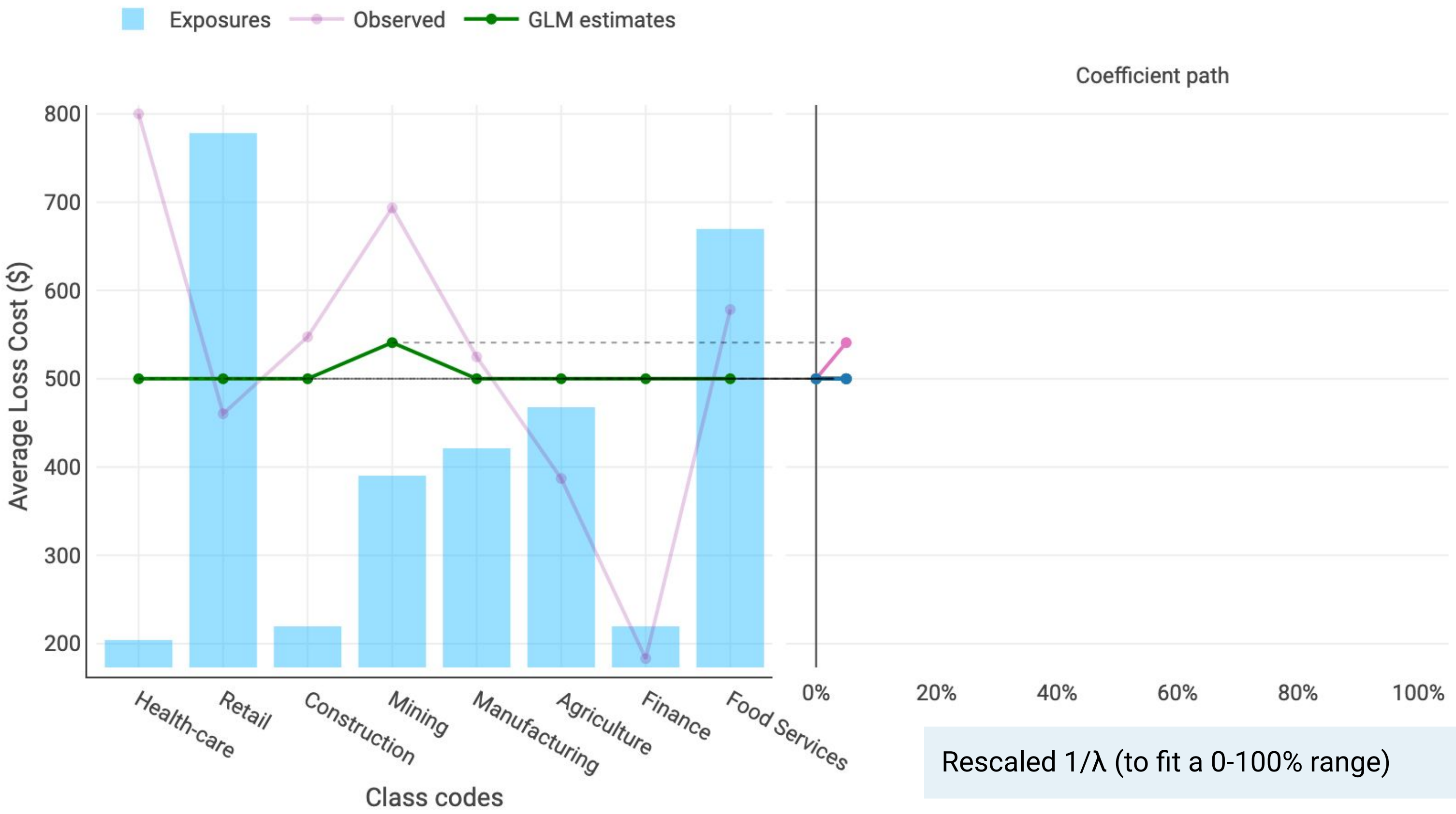


Coefficient path graph of Lasso



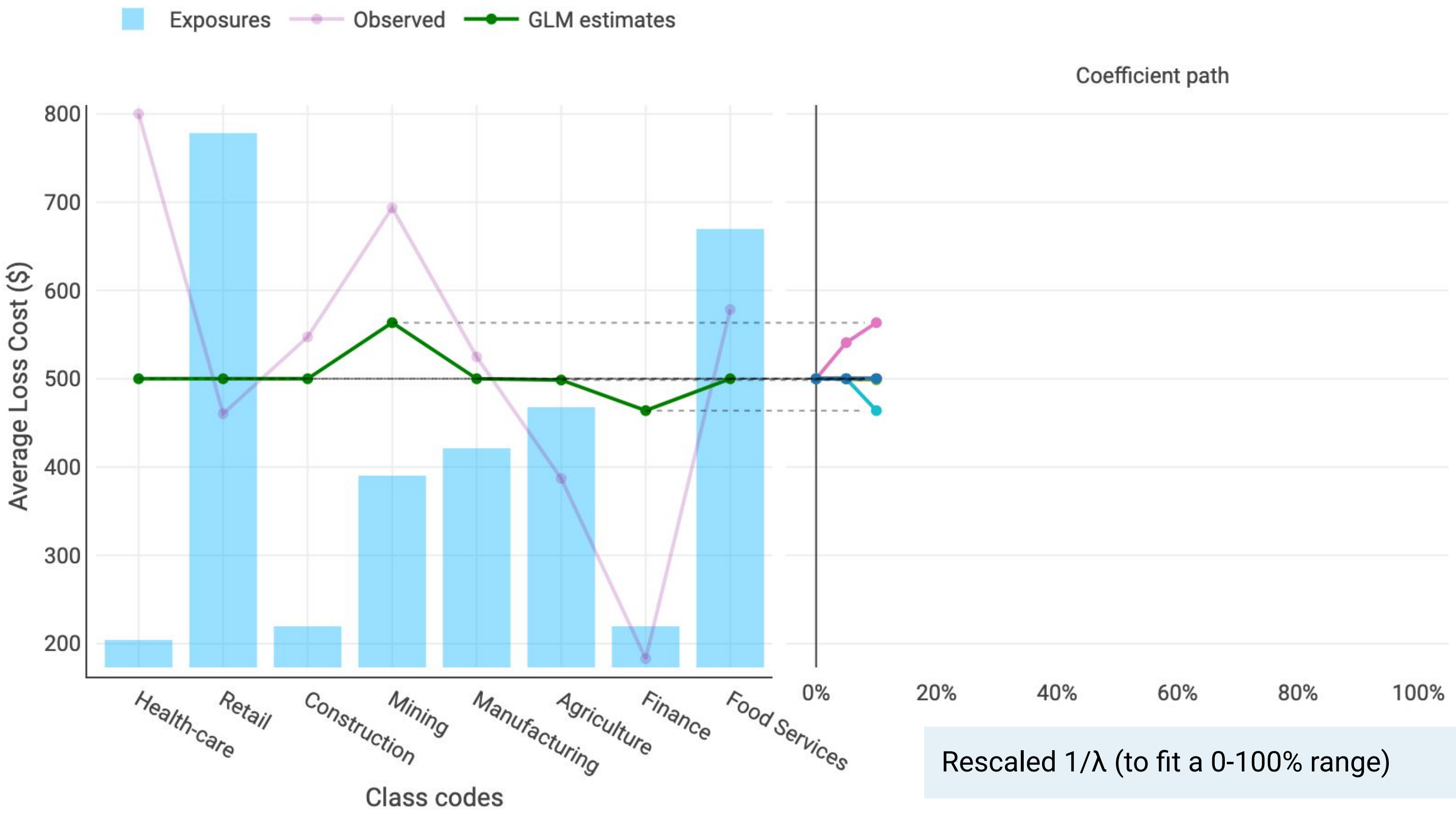
Rescaled 1/λ (to fit a 0-100% range)

Coefficient path graph of Lasso



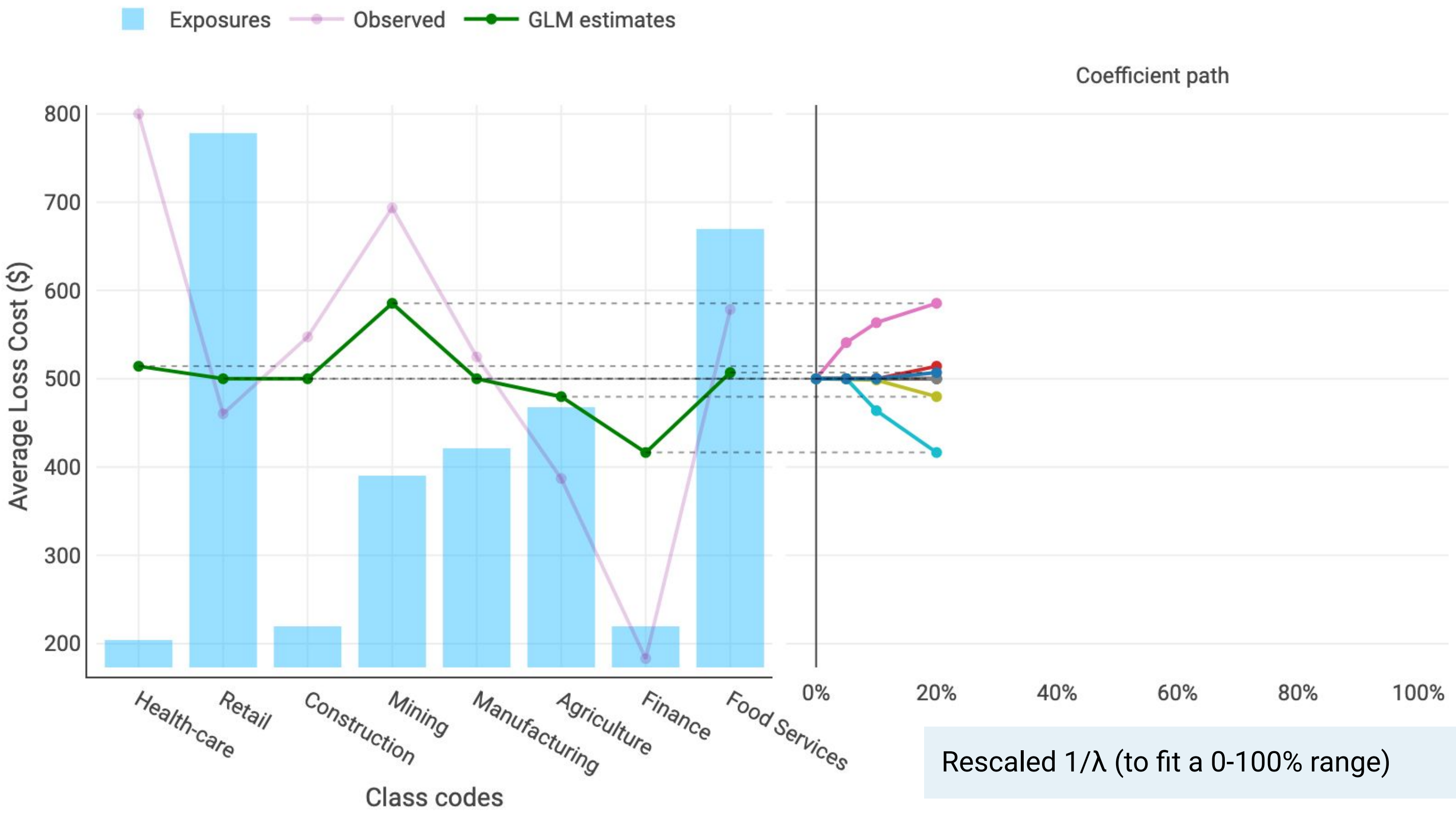
Rescaled 1/λ (to fit a 0-100% range)

Coefficient path graph of Lasso



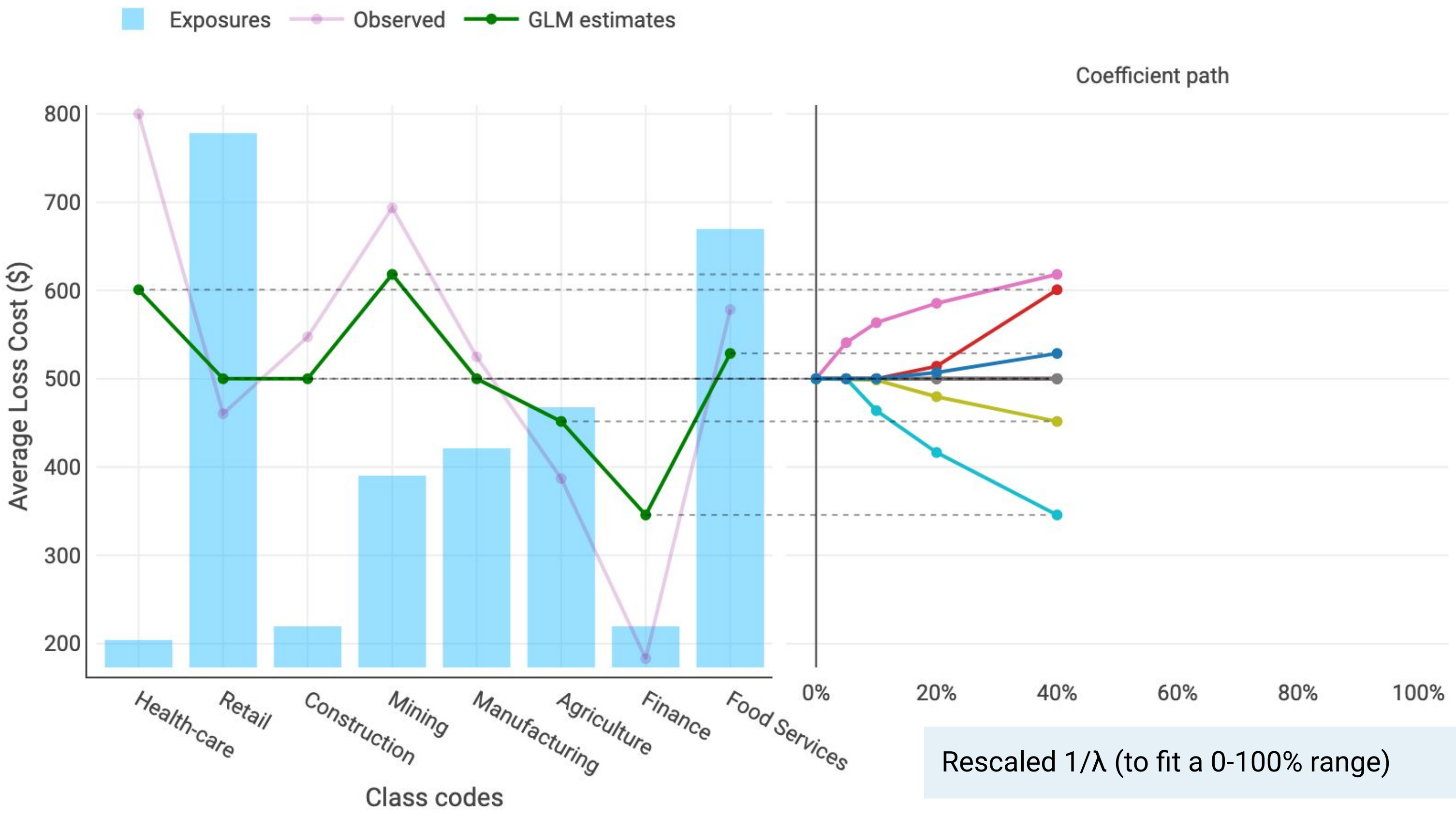
Rescaled 1/λ (to fit a 0-100% range)

Coefficient path graph of Lasso



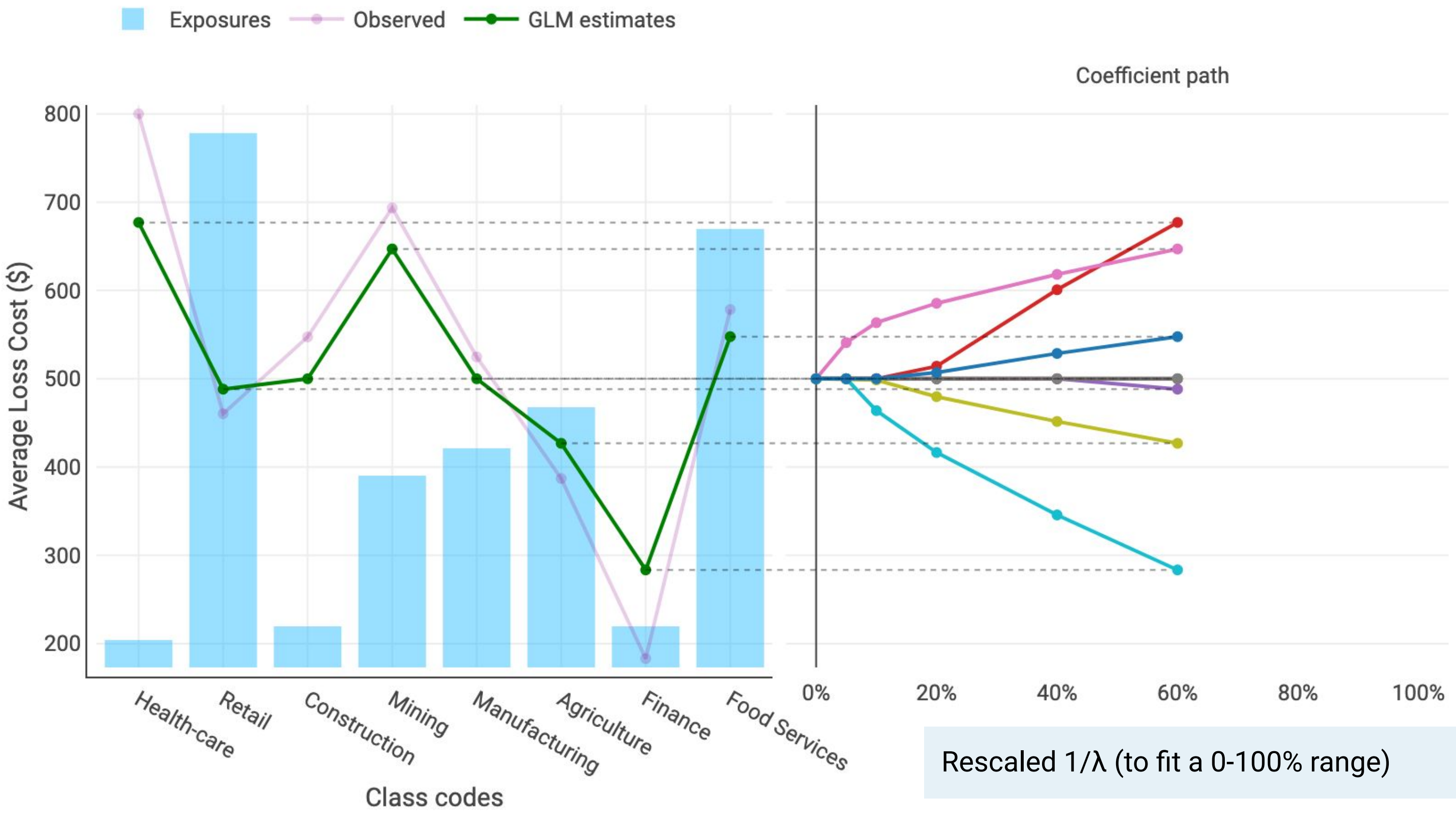
Rescaled 1/λ (to fit a 0-100% range)

Coefficient path graph of Lasso



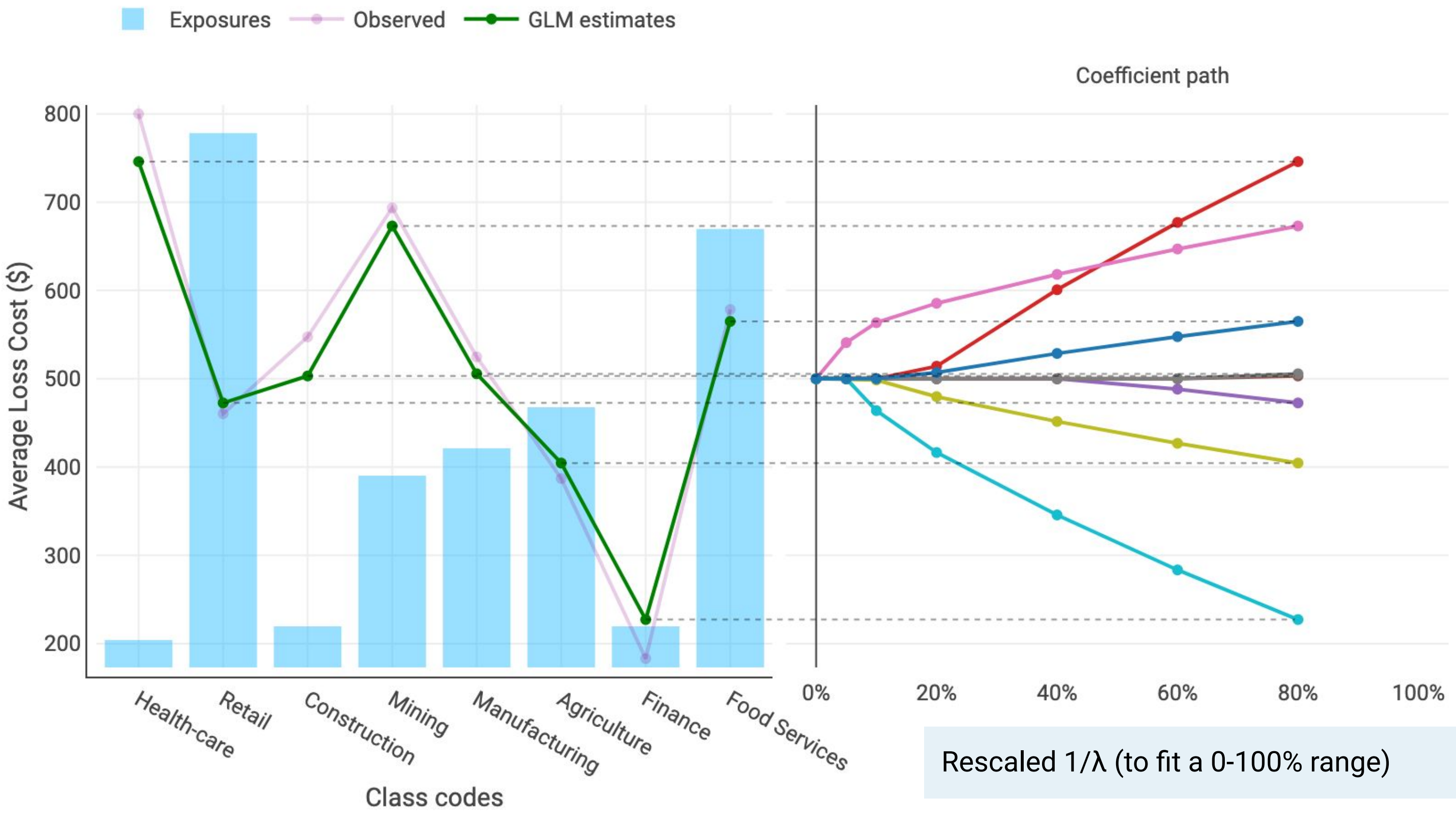
Rescaled 1/λ (to fit a 0-100% range)

Coefficient path graph of Lasso



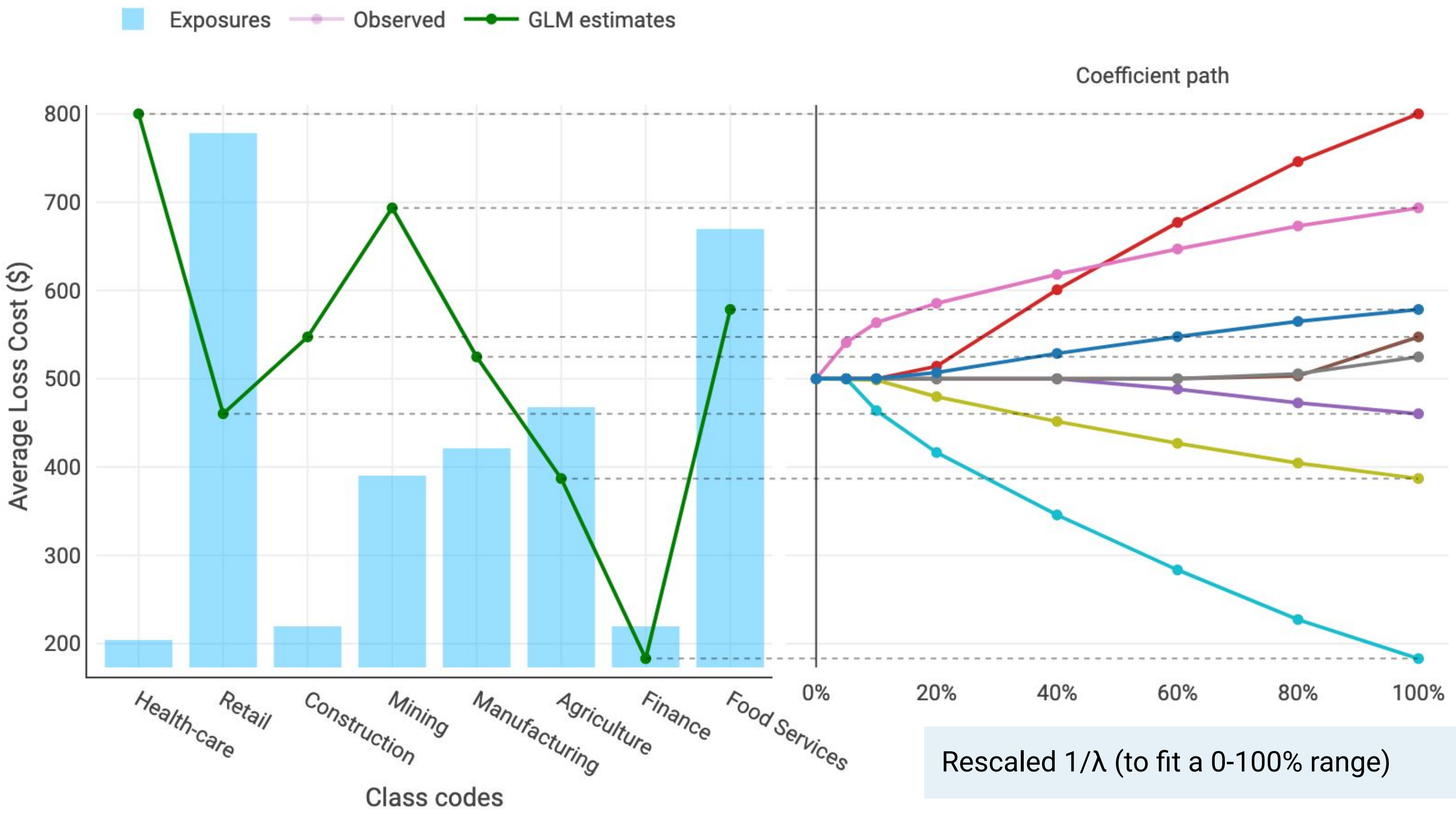
Rescaled 1/λ (to fit a 0-100% range)

Coefficient path graph of Lasso



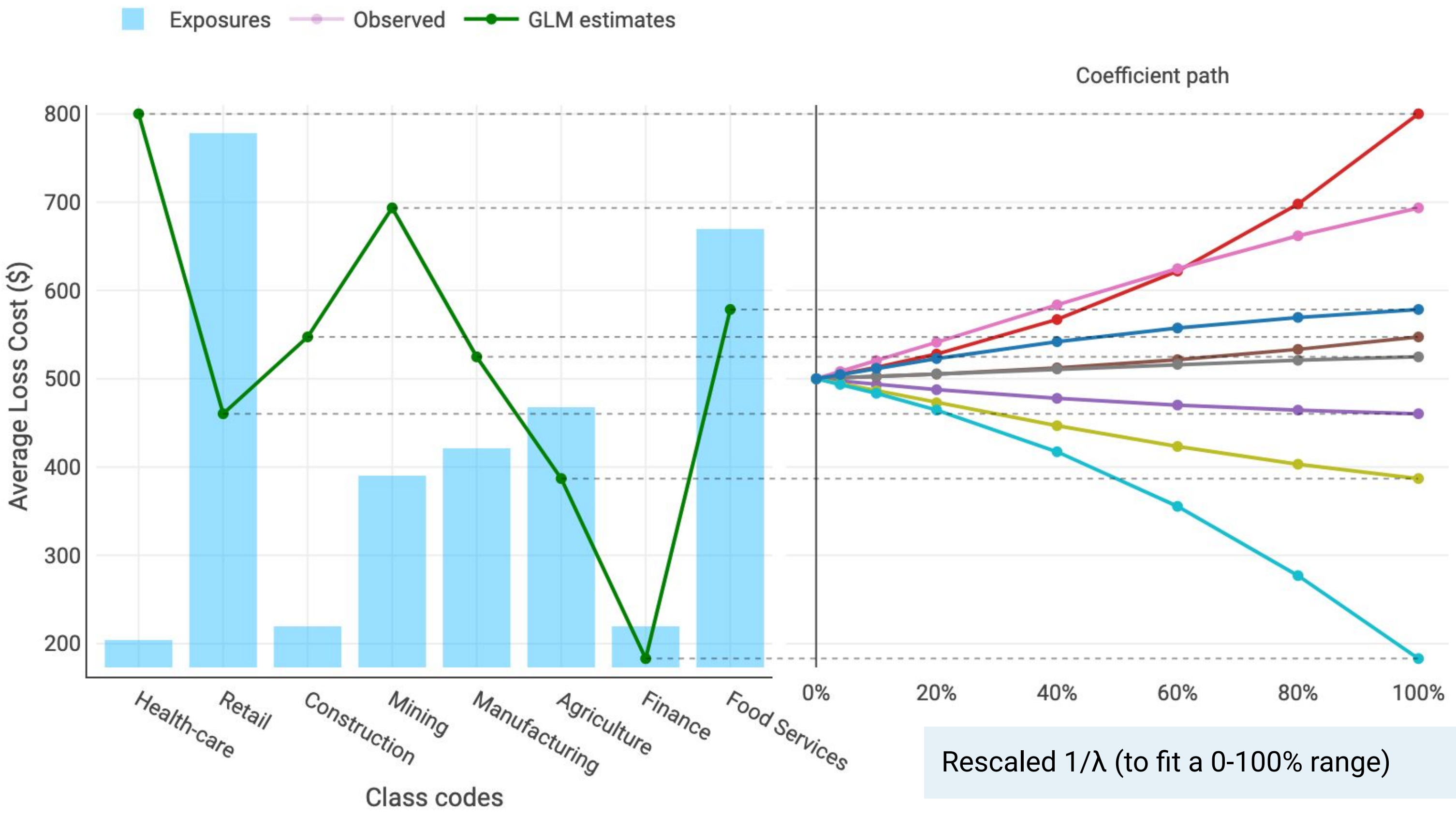
Rescaled 1/λ (to fit a 0-100% range)

Coefficient path graph of Lasso



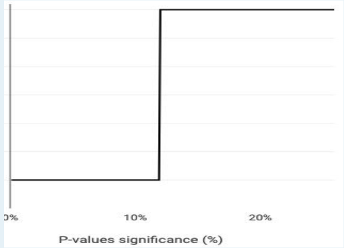
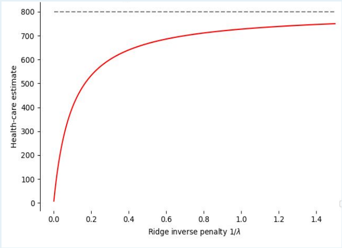
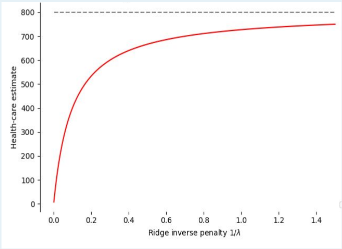
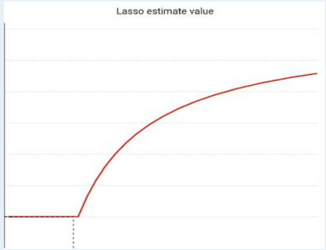
Rescaled 1/λ (to fit a 0-100% range)

Coefficient path graph of Ridge



Rescaled 1/λ (to fit a 0-100% range)

Comparing Different techniques

	GLM Significance	Buhlmann Credibility	Ridge	Lasso
Inclusion Threshold	Yes	No	No	Yes
Credibility Weighting	No	Yes	Yes	Yes
Multivariate	Yes	No	Yes	Yes
Credibility Assignment				

Innovations in Penalization

AKUR8

Can we let penalization create our models for us?

AKUR8

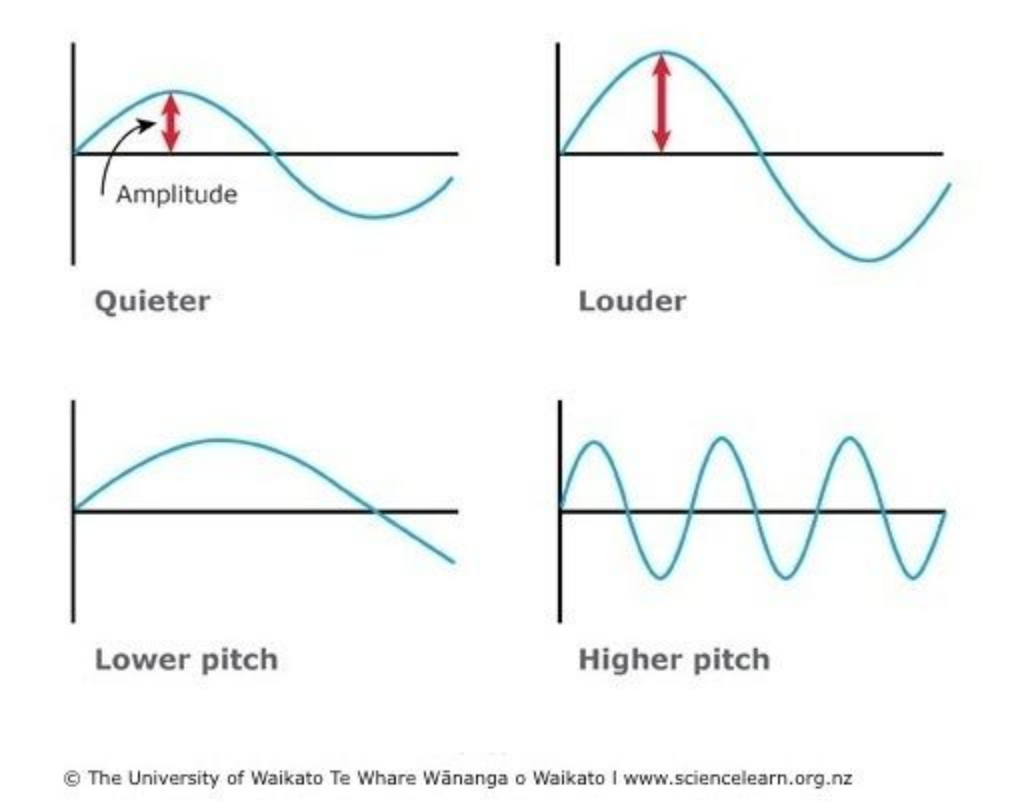
Musical Interlude

AKUR8

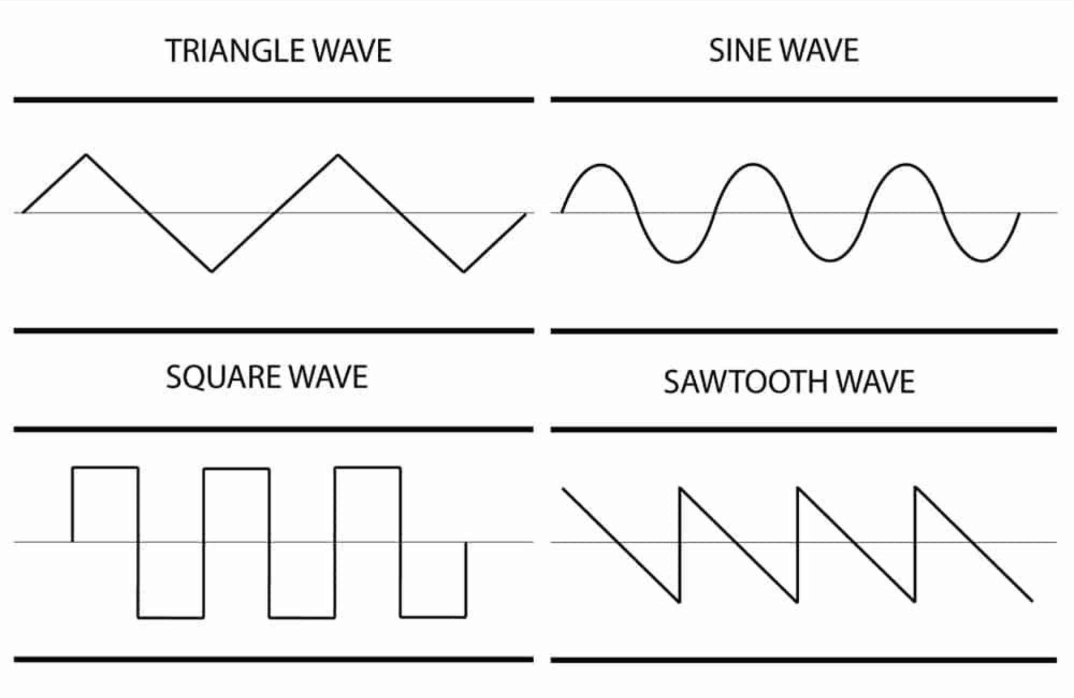
What is Sound?

AKUR8

A Simple Sound is a Sine Wave

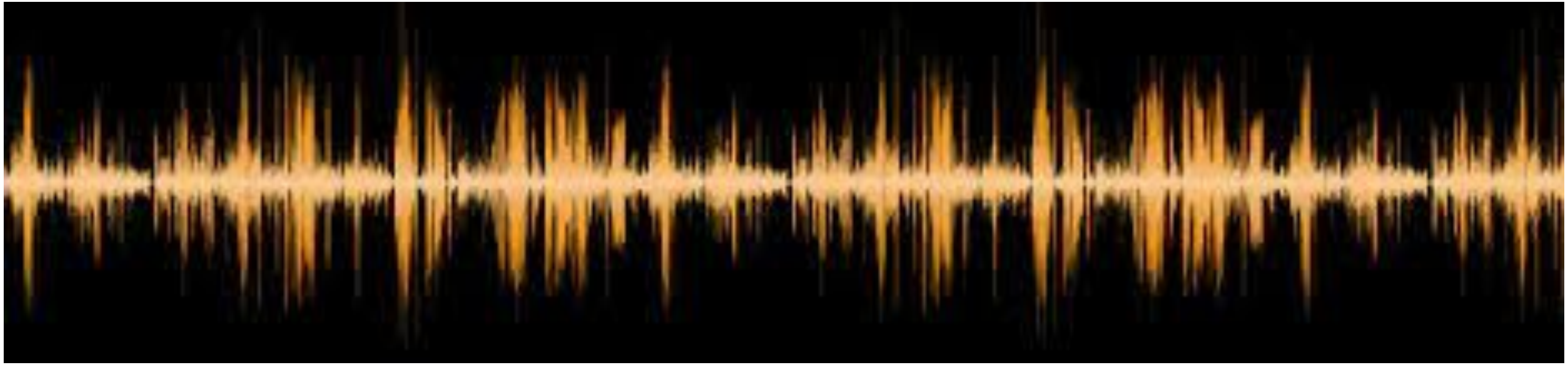


Sound Waves can be More Complex



https://primesound.org/how-do-synthesizers-work/

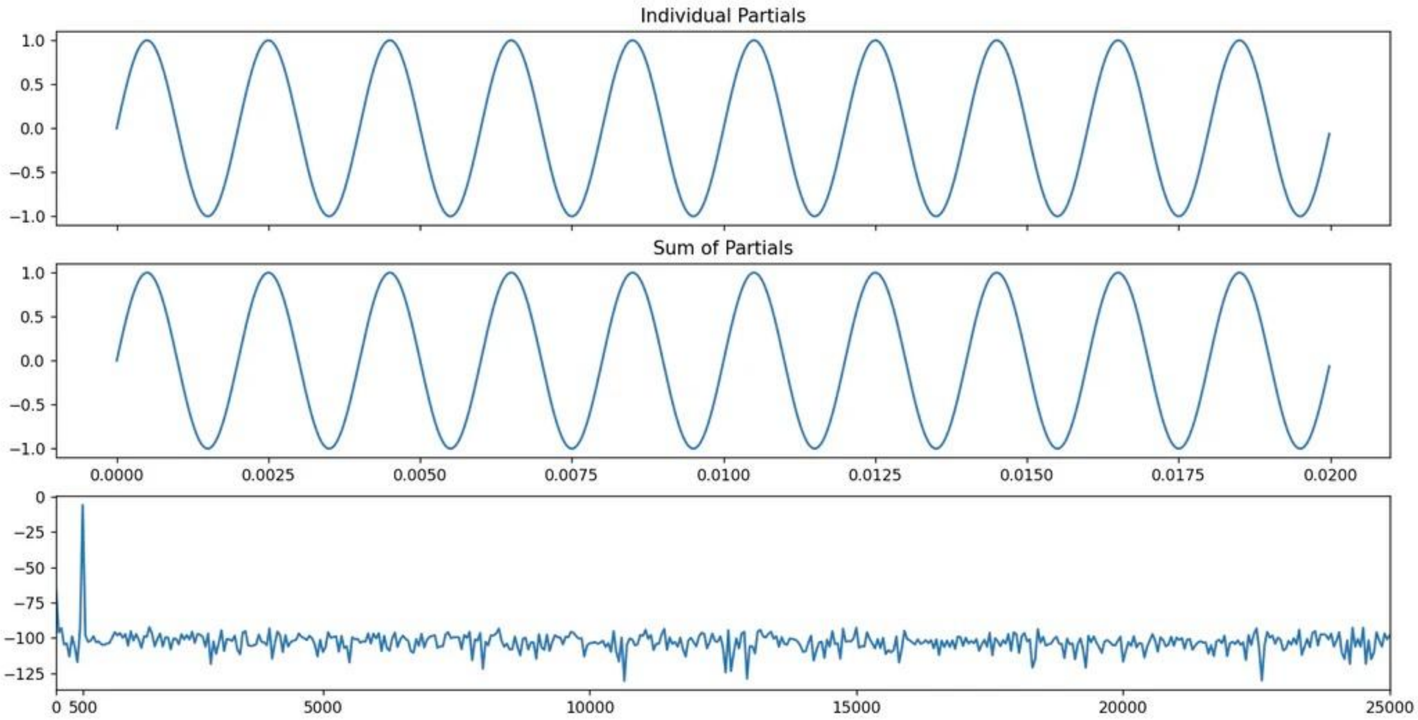
Sound Waves can be Very Complex



How does a Synthesizer create Sounds?

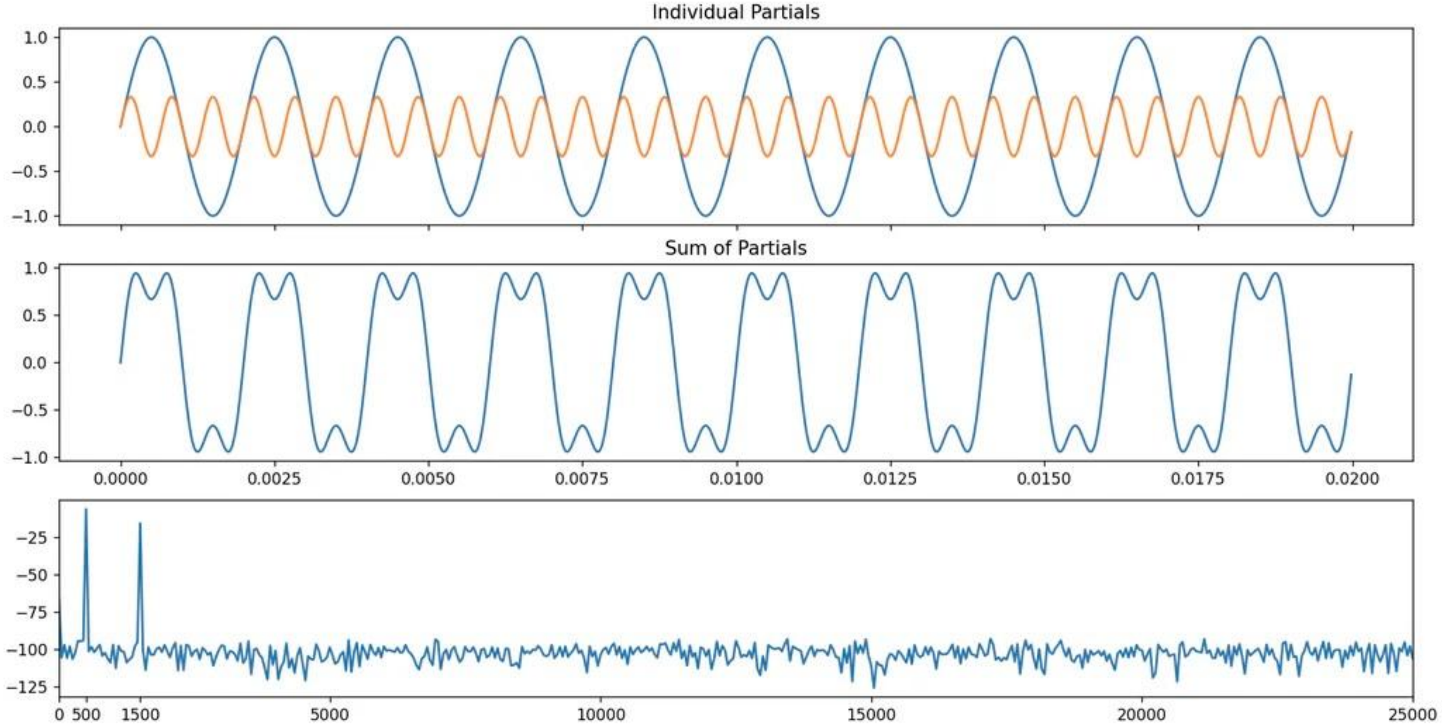
https://differentialaudioinc.com/blogs/news/intro-to-additive-synthesis

Synthesizers use "Additive Synthesis" to create Complex Sounds



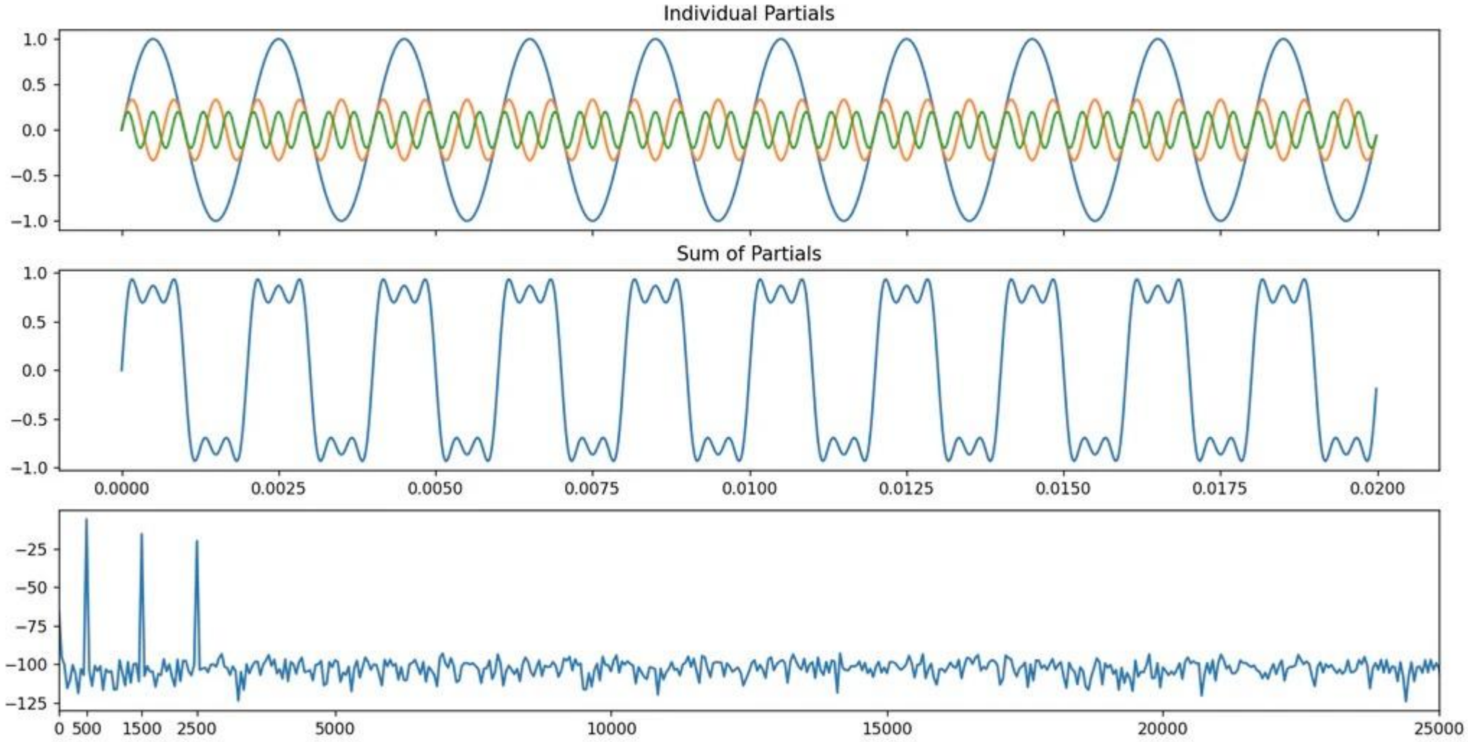
https://differentialaudioinc.com/blogs/news/intro-to-additive-synthesis

Synthesizers use “Additive Synthesis” to create Complex Sounds



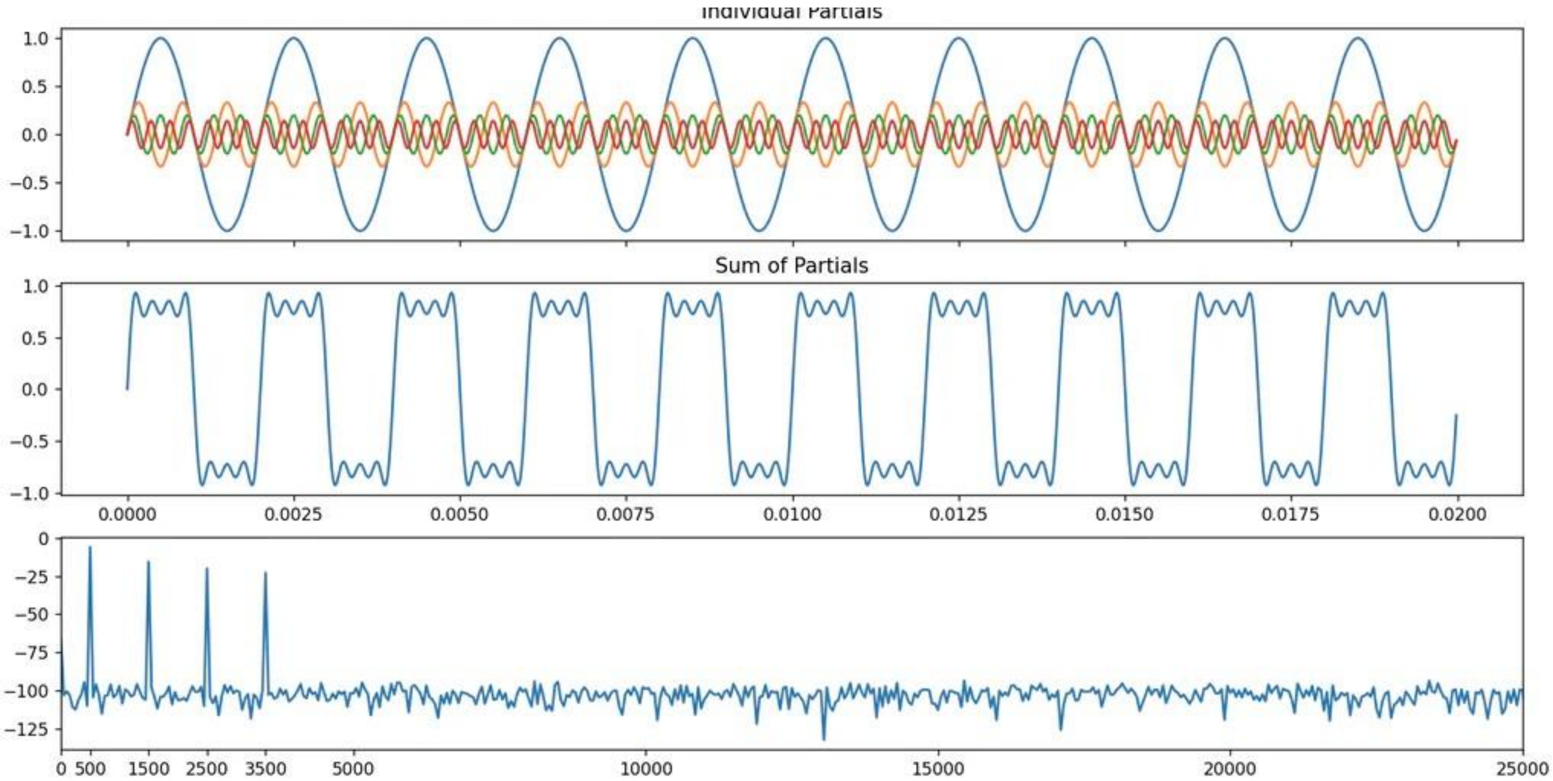
https://differentialaudioinc.com/blogs/news/intro-to-additive-synthesis

Synthesizers use “Additive Synthesis” to create Complex Sounds



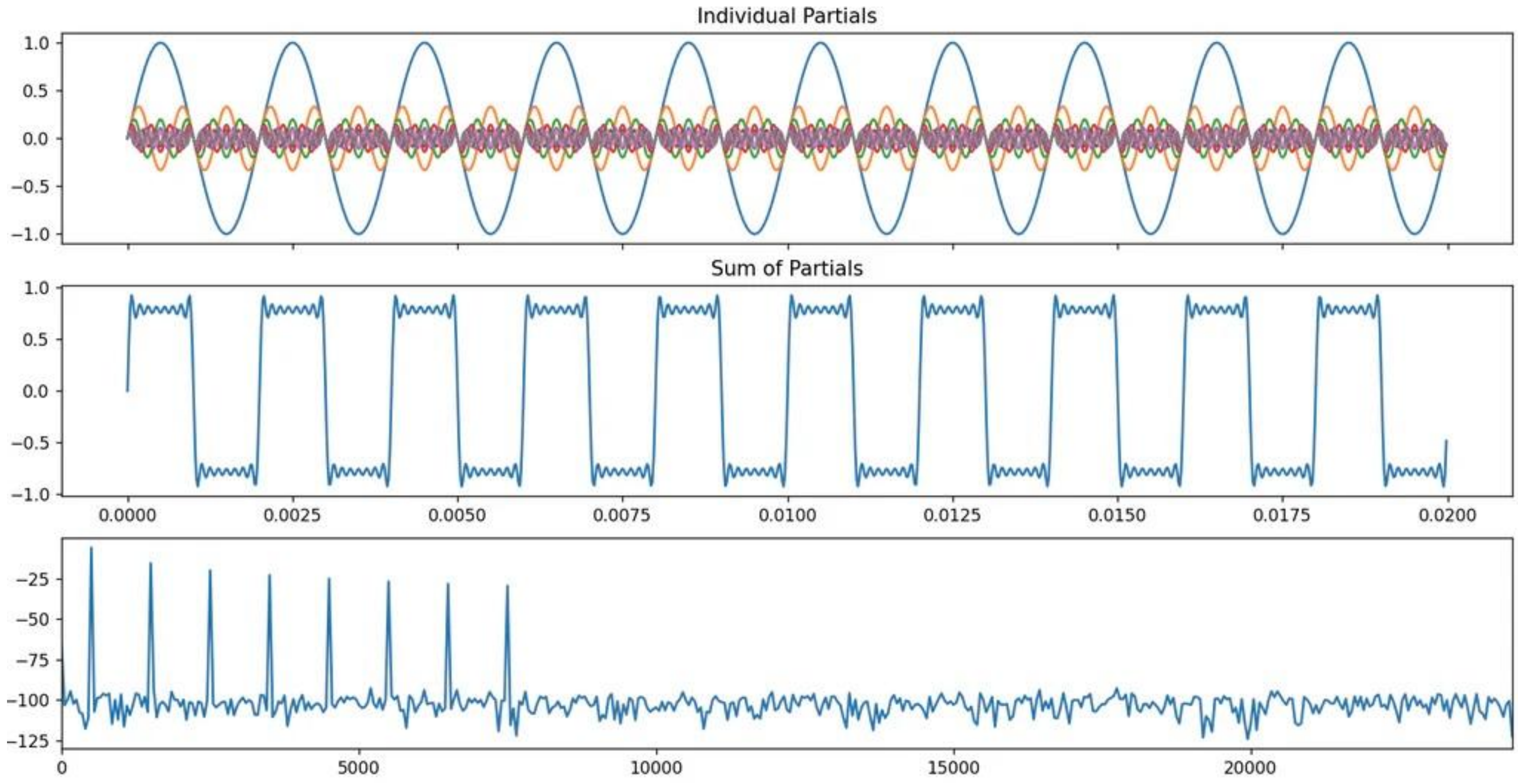
https://differentialaudioinc.com/blogs/news/intro-to-additive-synthesis

Synthesizers use “Additive Synthesis” to create Complex Sounds



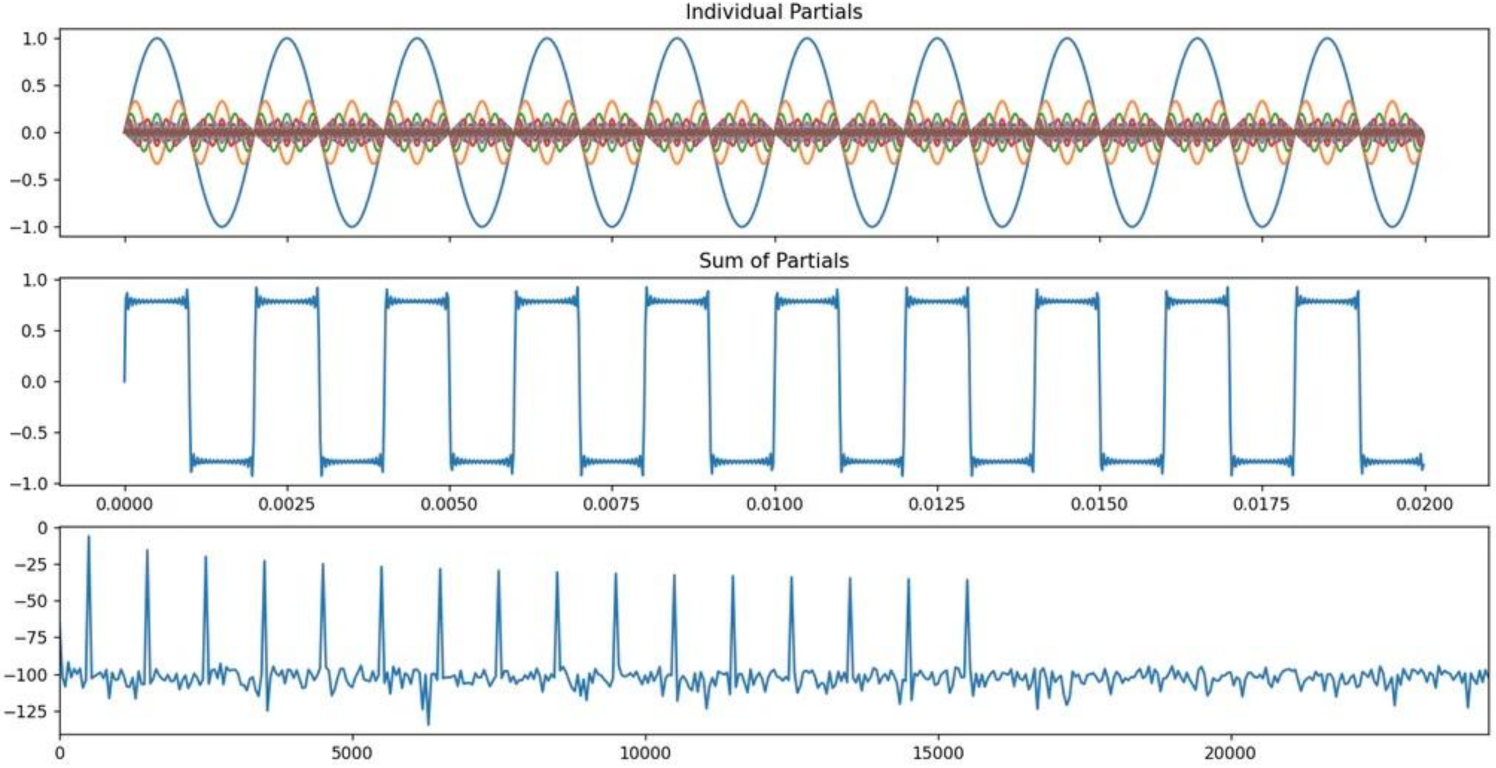
https://differentialaudioinc.com/blogs/news/intro-to-additive-synthesis

Synthesizers use “Additive Synthesis” to create Complex Sounds



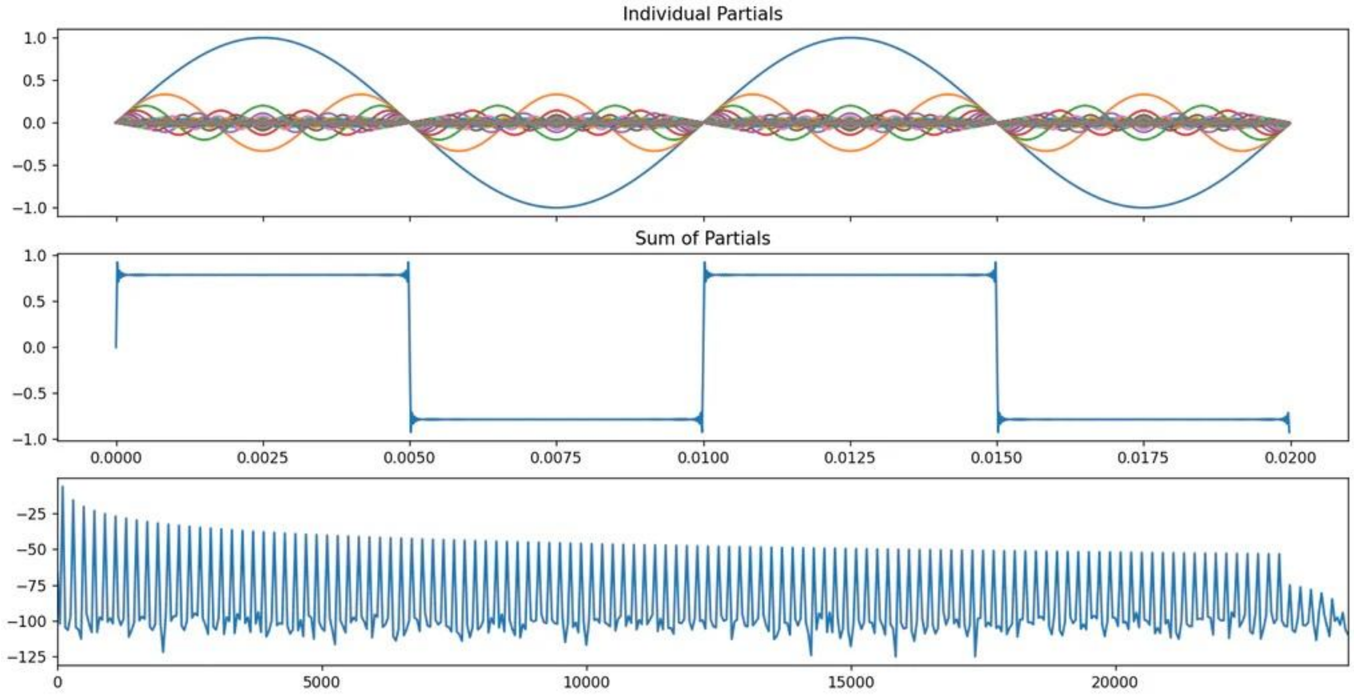
https://differentialaudioinc.com/blogs/news/intro-to-additive-synthesis

Synthesizers use “Additive Synthesis” to create Complex Sounds



https://differentialaudioinc.com/blogs/news/intro-to-additive-synthesis

Synthesizers use “Additive Synthesis” to create Complex Sounds



Additive Synthesis creates something complex out of something basic

AKUR8

GAMs are Synthesizers for Statistics

AKUR8

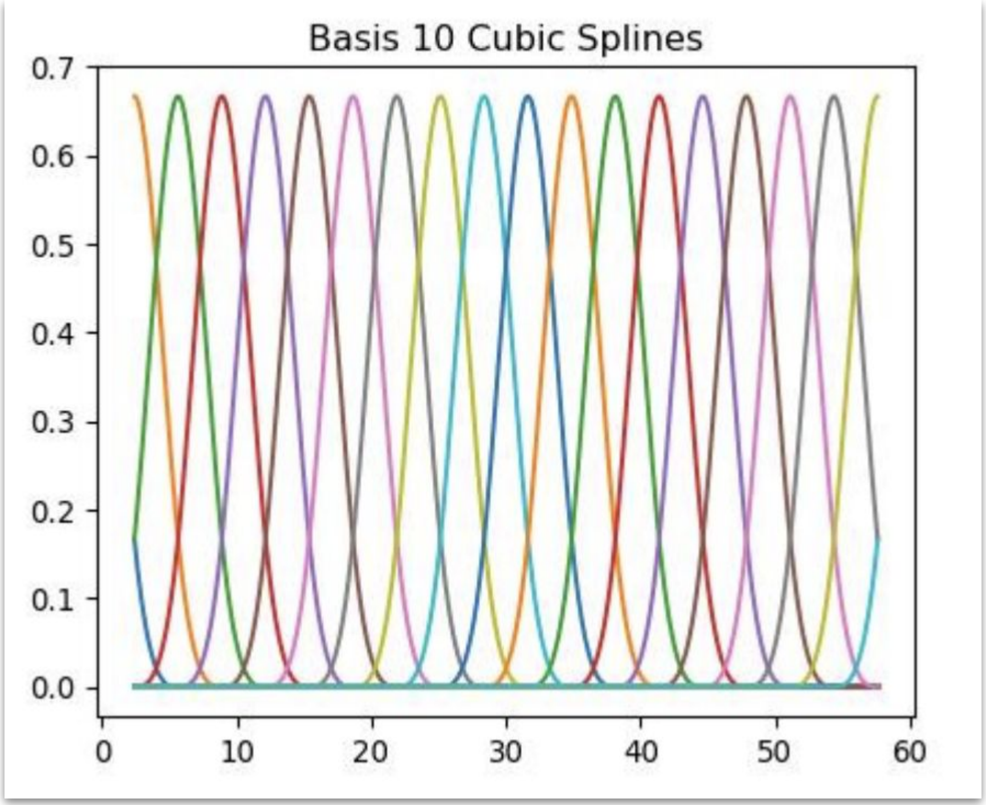
Generalized **Additive** Models are **Additive** Synthesis for Statistics

AKUR8

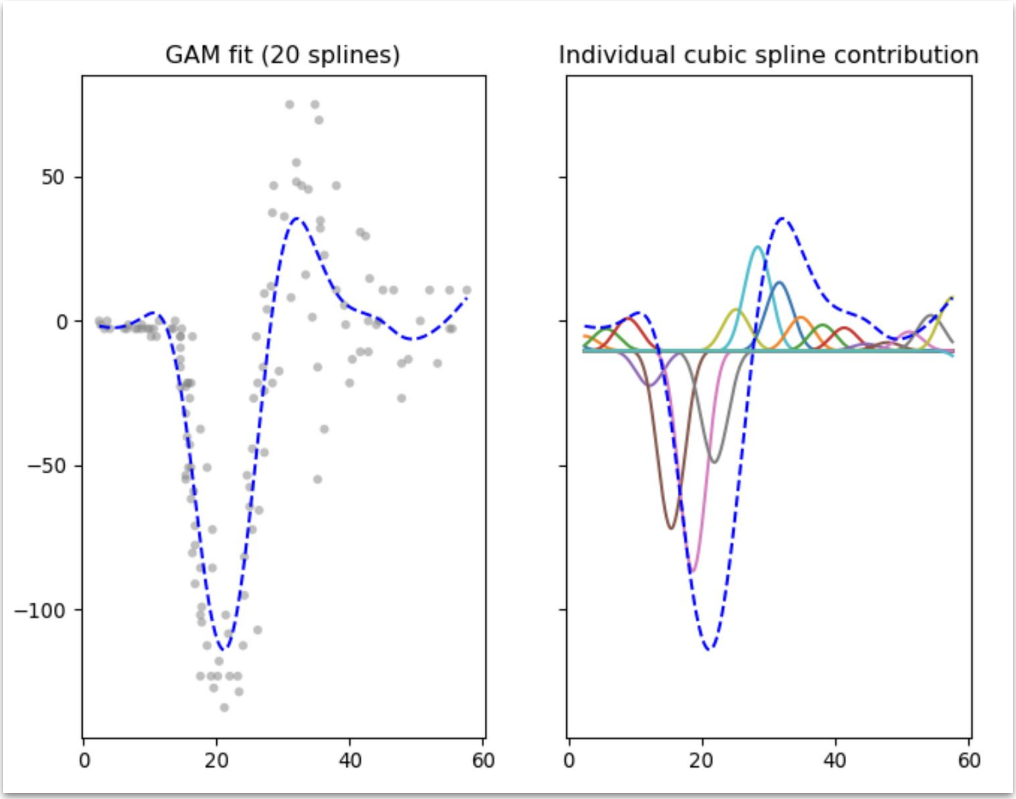
Penalization Driven Models

Traditional “MGCV” GAM

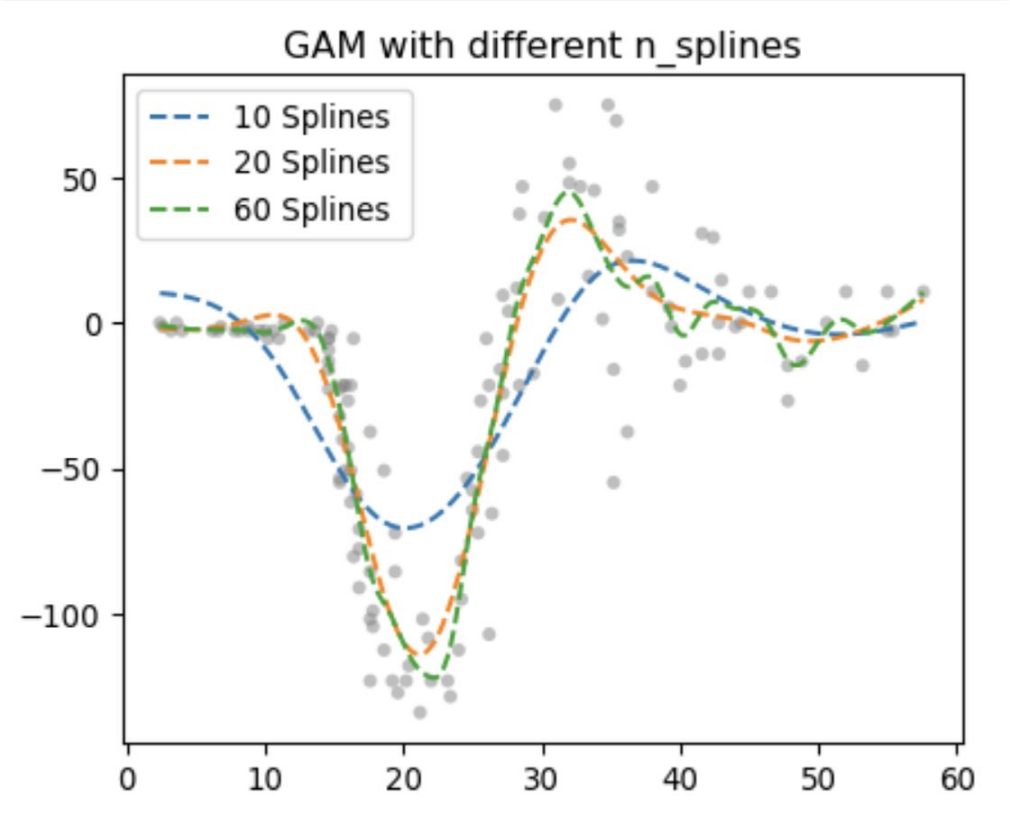
A Good Basis Function can Fit to Complex Data



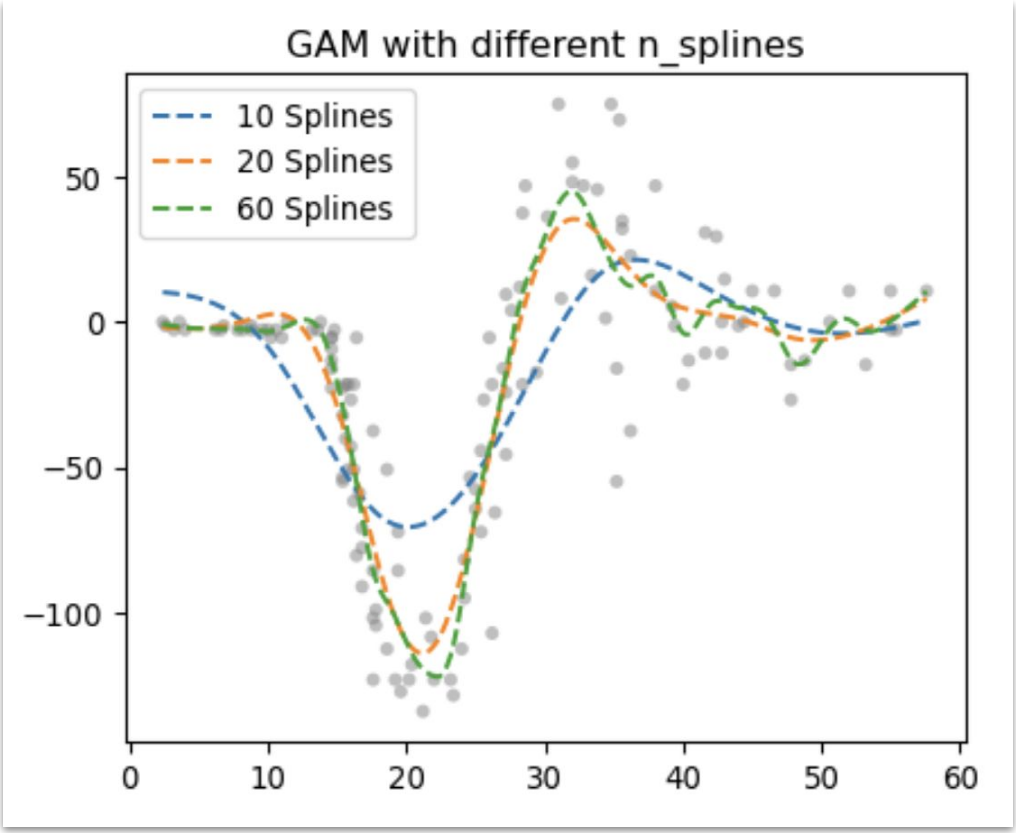
Example MGCV GAM Model Fit



Example MGCV GAM Model Fit



Avoiding Overfitting in MGCV GAM



High **concurvity** occurs when one of the cubic spline terms can be approximated by using the other cubic spline terms. Fewer splines should be selected.

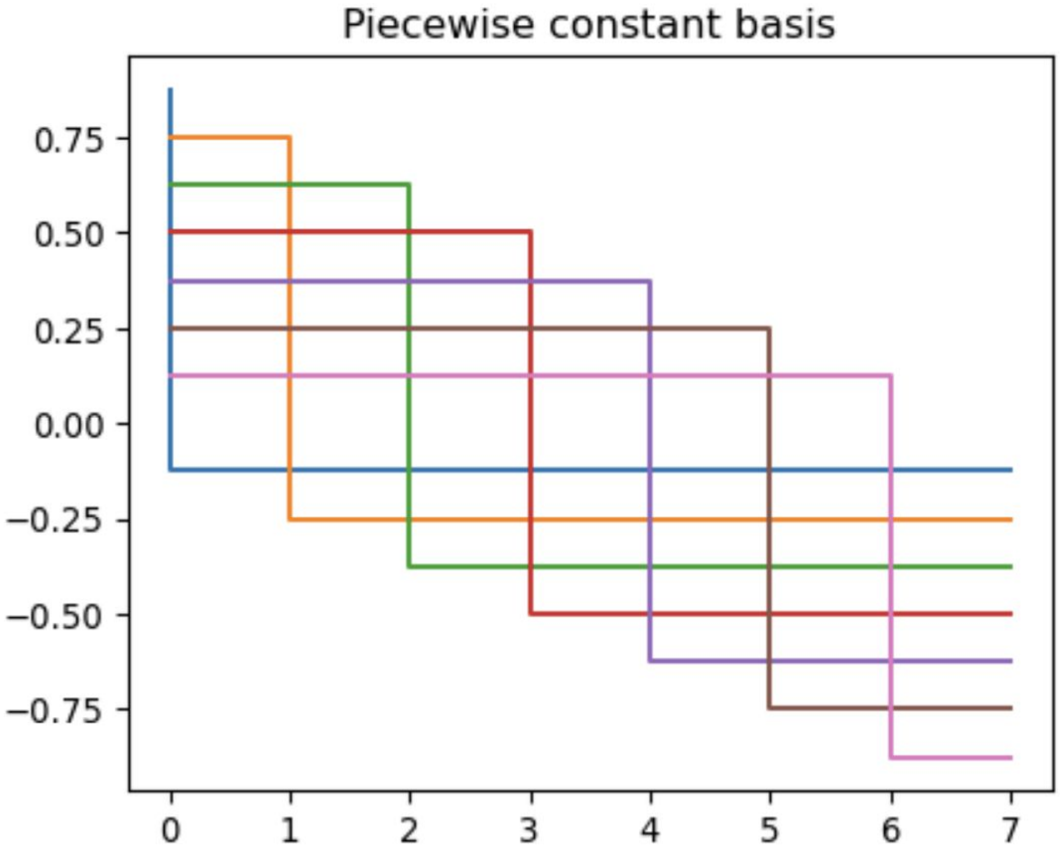
Ridge **penalization** can be applied to provide stability and avoid overfitting.

Penalization determines the amount of each basic element used to create the complex shape

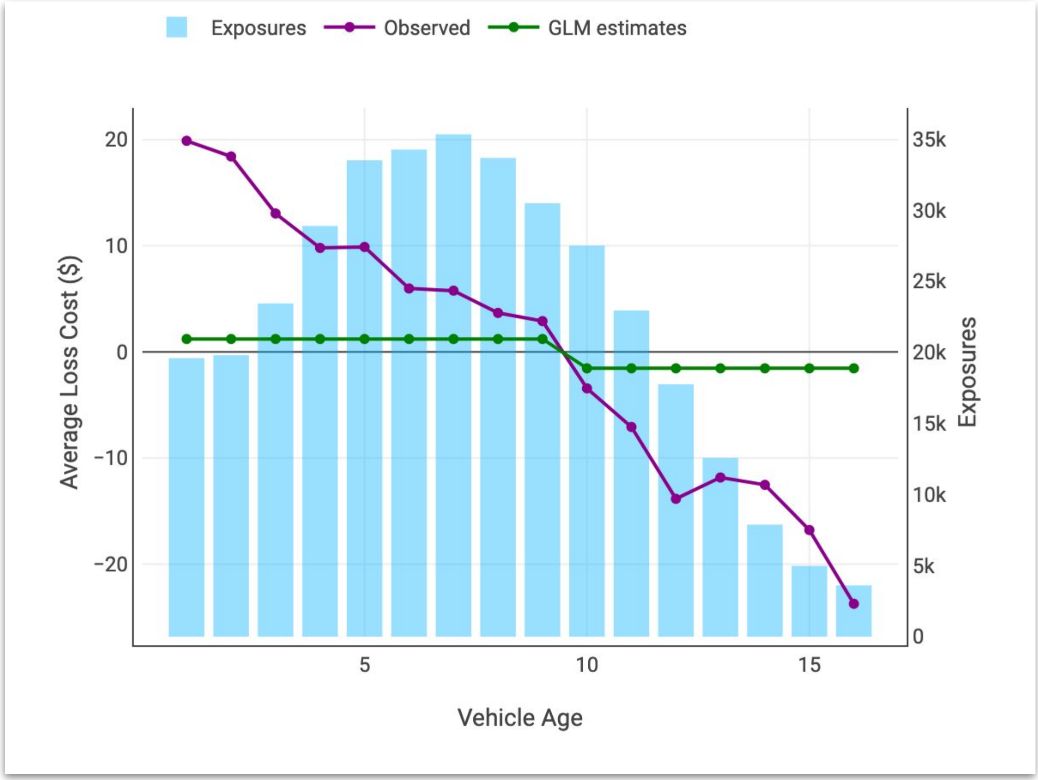
AKUR8

Derivative Lasso

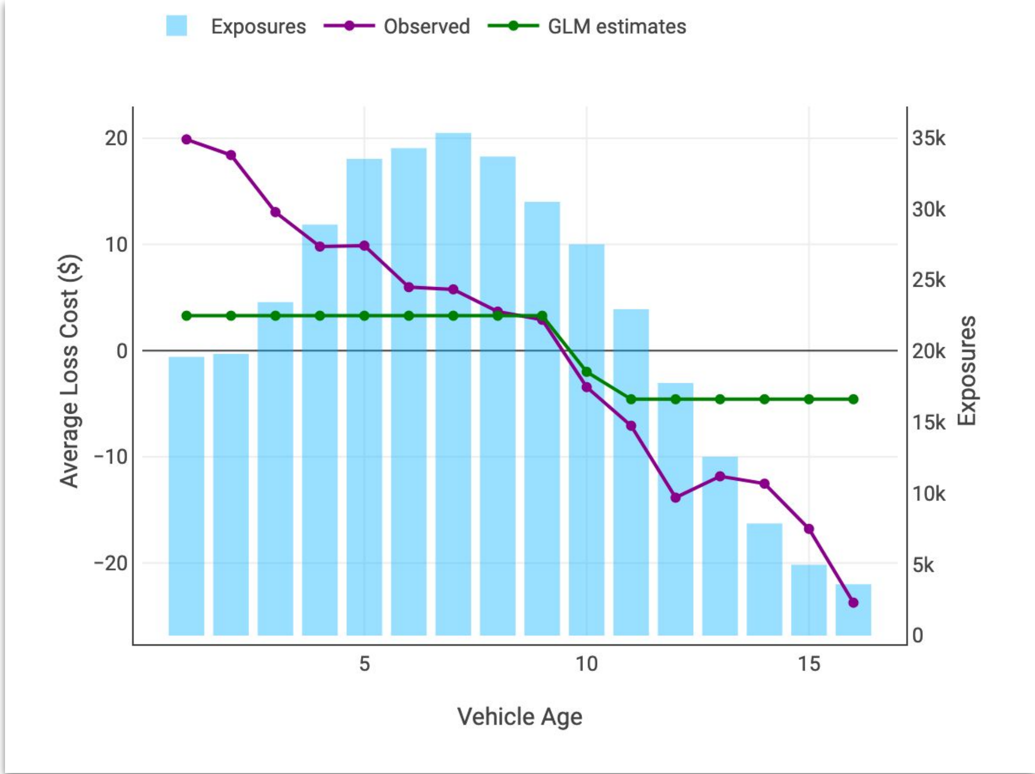
A Good Basis Function can Fit to Complex Data



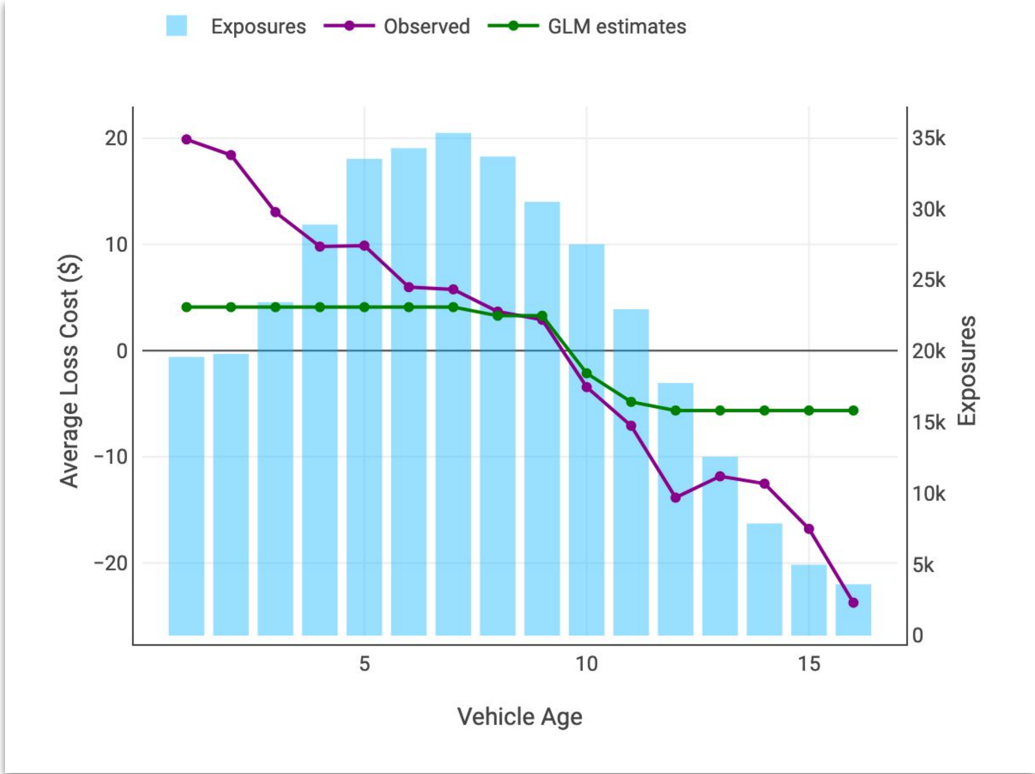
Derivative Lasso and Ordinal variables



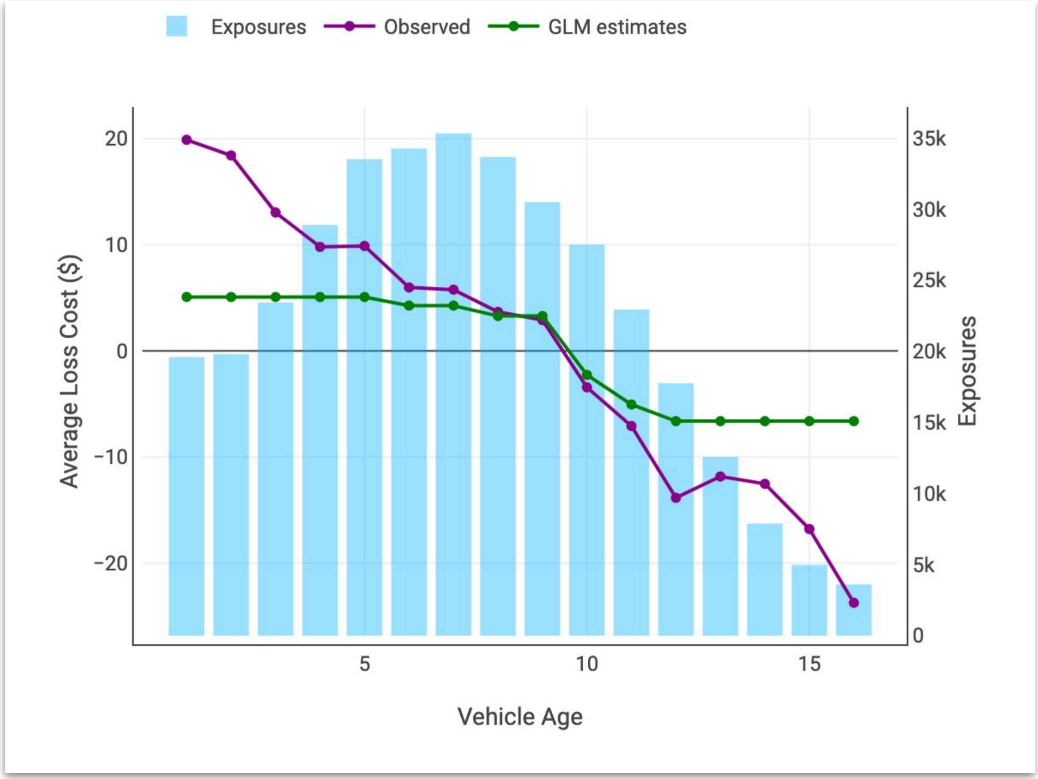
Derivative Lasso and Ordinal variables



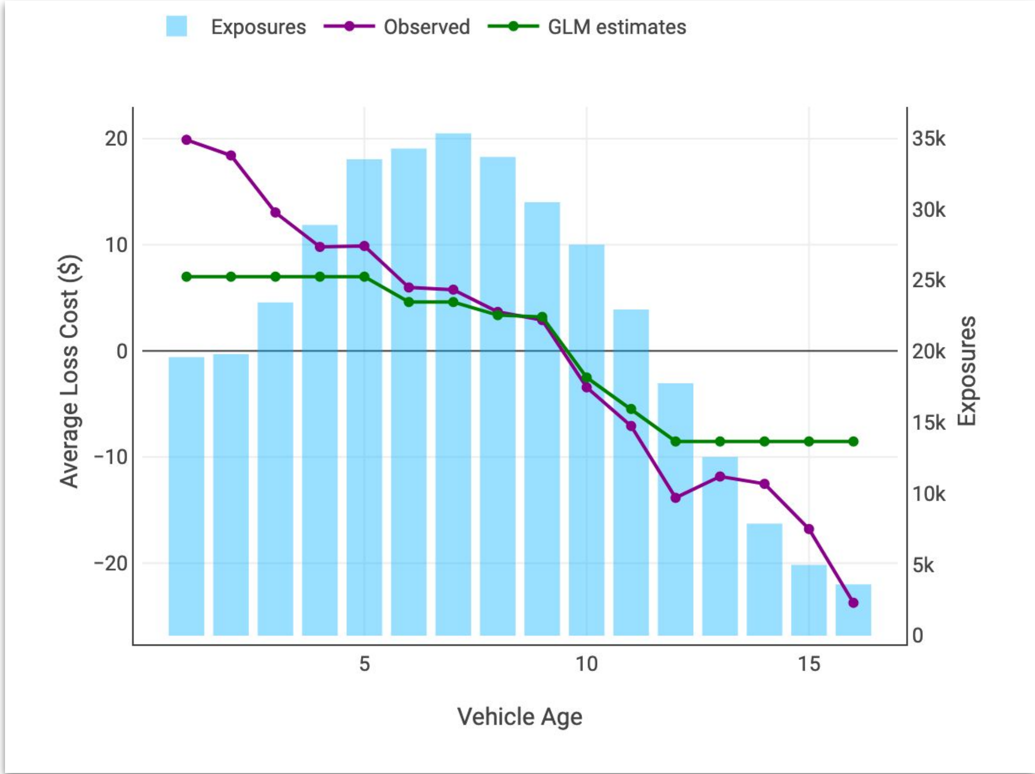
Derivative Lasso and Ordinal variables



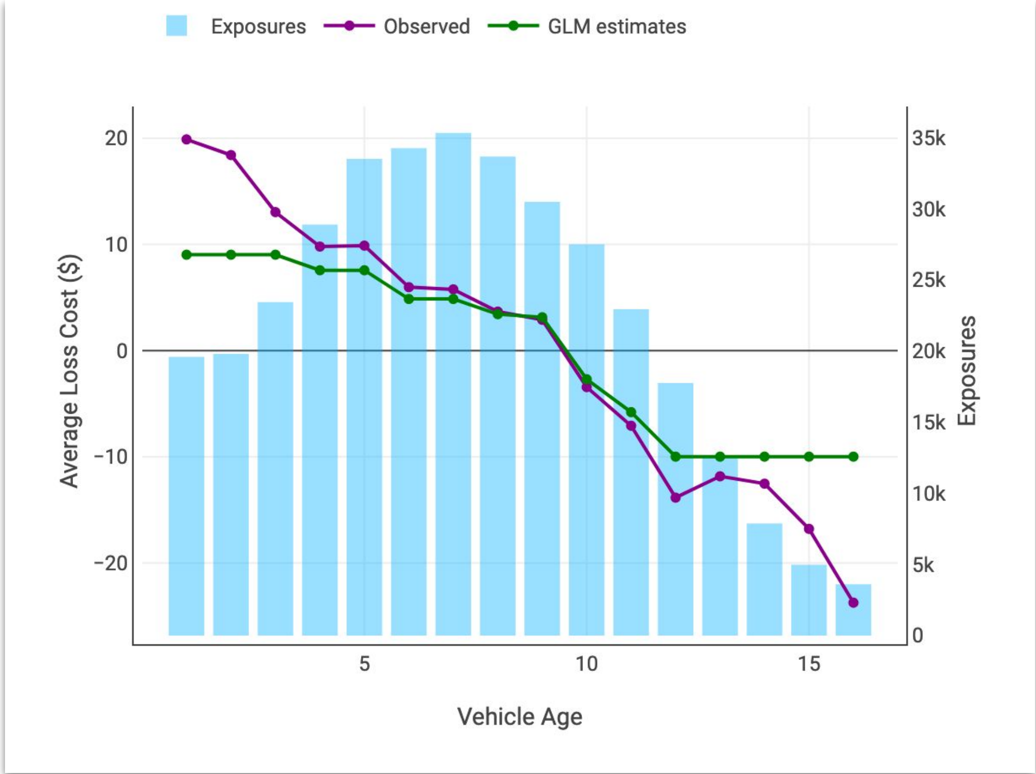
Derivative Lasso and Ordinal variables



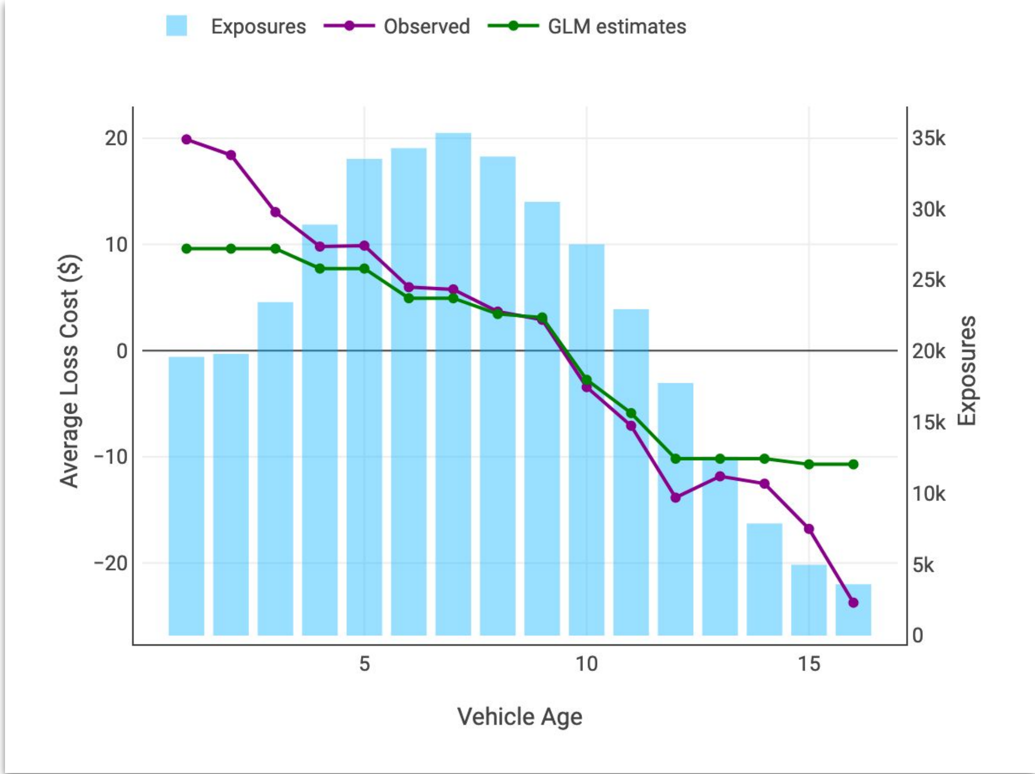
Derivative Lasso and Ordinal variables



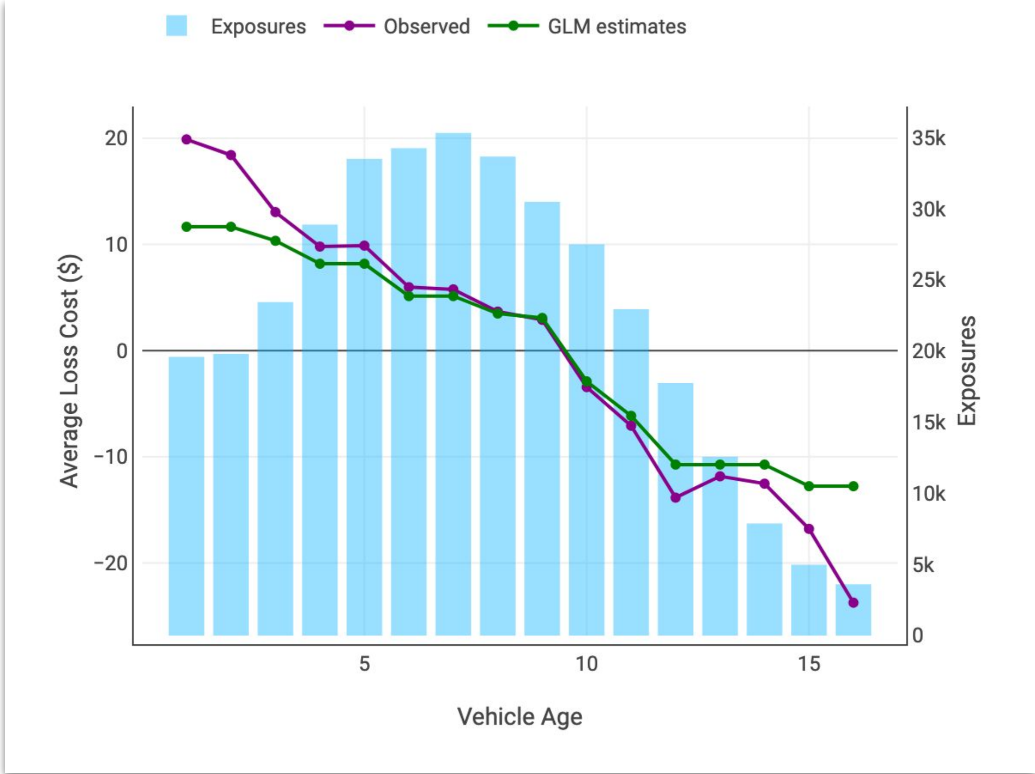
Derivative Lasso and Ordinal variables



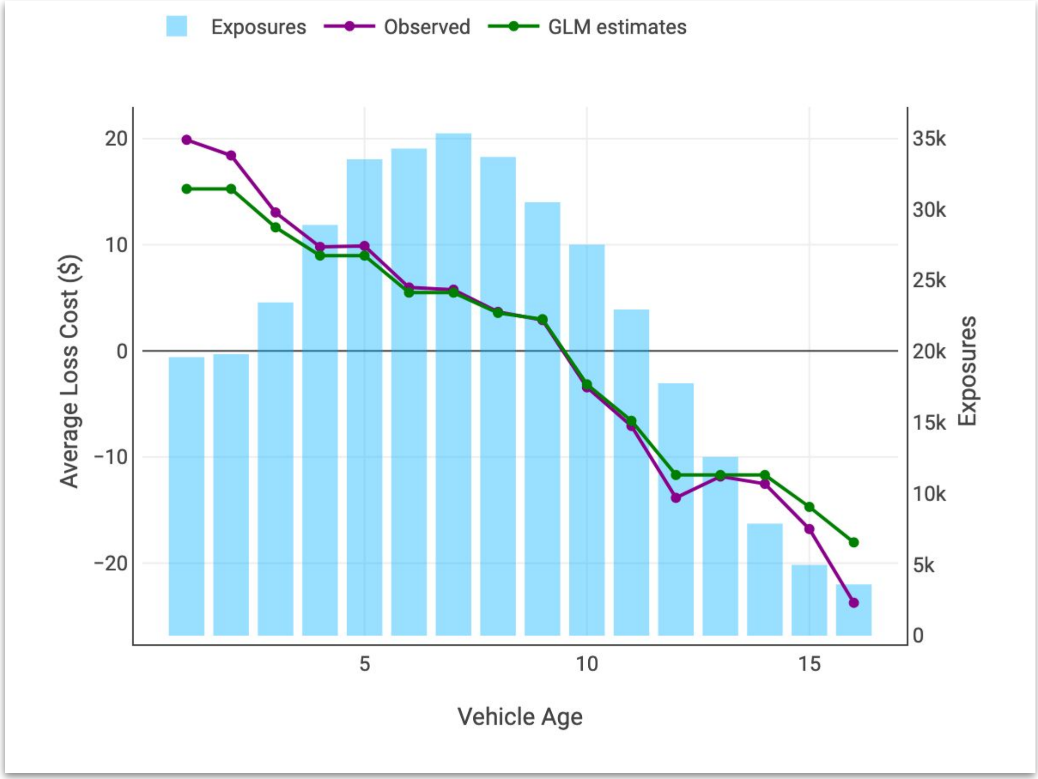
Derivative Lasso and Ordinal variables



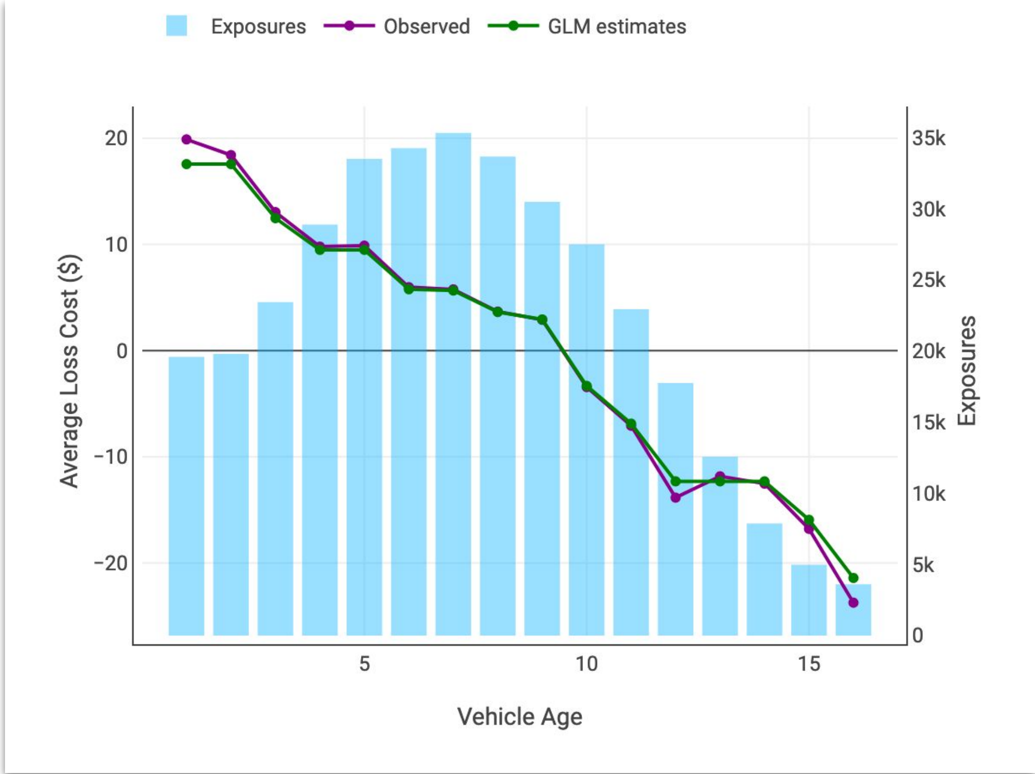
Derivative Lasso and Ordinal variables



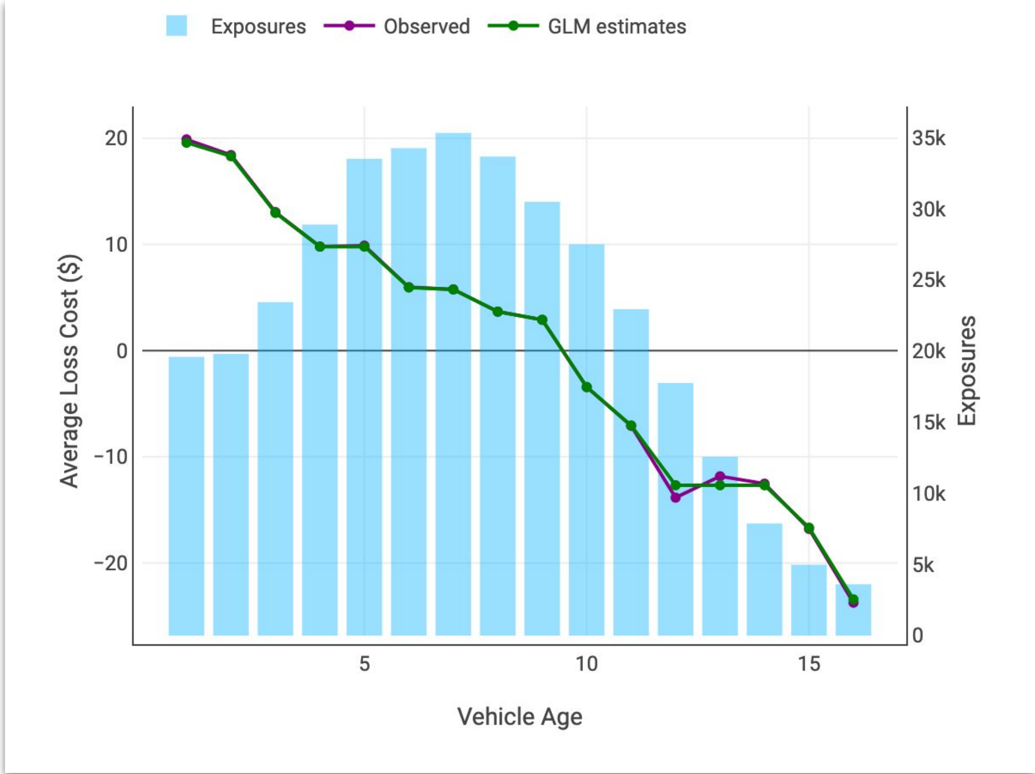
Derivative Lasso and Ordinal variables



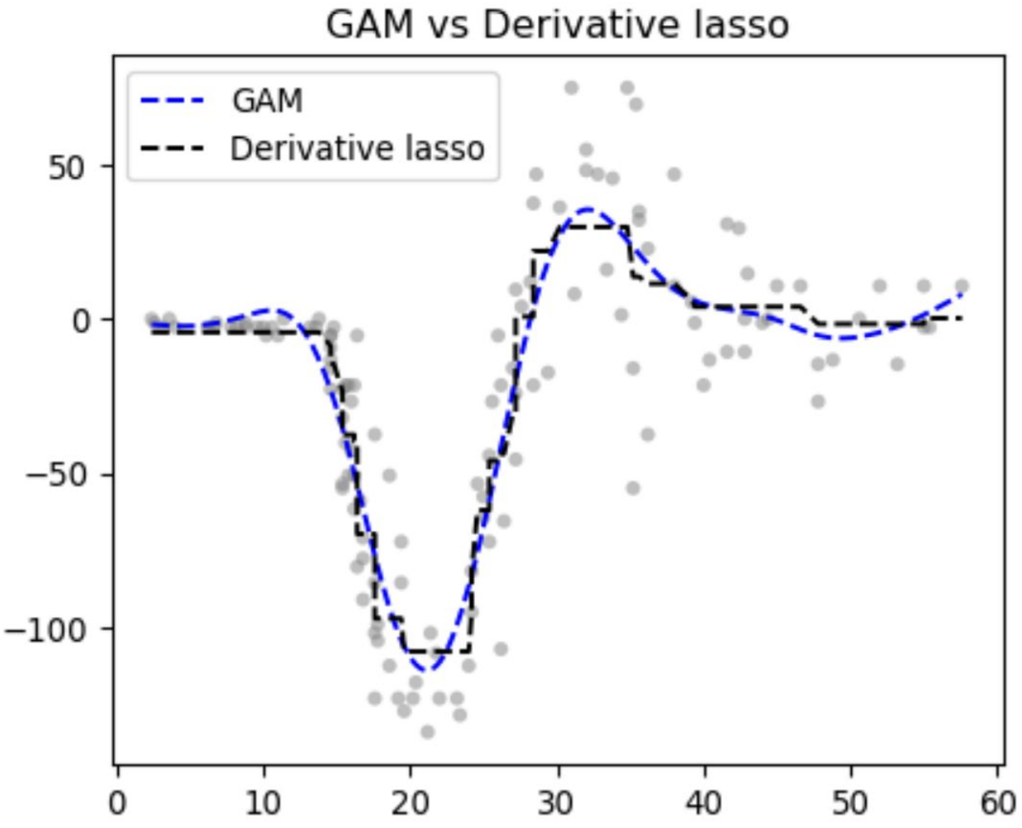
Derivative Lasso and Ordinal variables



Derivative Lasso and Ordinal variables



Derivative Lasso vs. MGCV GAM



You haven't mentioned the penalty term yet...

Actuarial Models are not Strictly Statistical!

1 **Predictive performance is not the only goal of a good model**

2 **Actuarial judgement / selections are a fundamental component of modeling**

The selection of a penalty term is strictly a **statistical** process for **most models.**

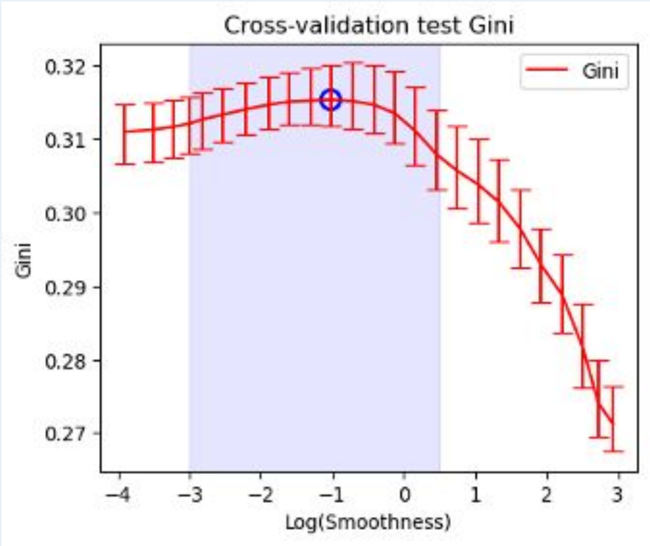
The selection of a penalty is both **statistical and actuarial** for **derivative lasso.**

Cross Validation: Selecting the Penalty Term

Performance:
vertical y axis

The higher - the better

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



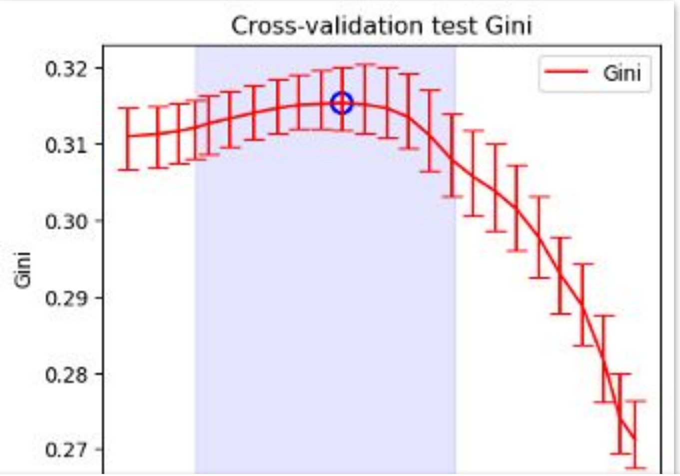
Penalty is evaluated from **low (equal to GLM) to high**.

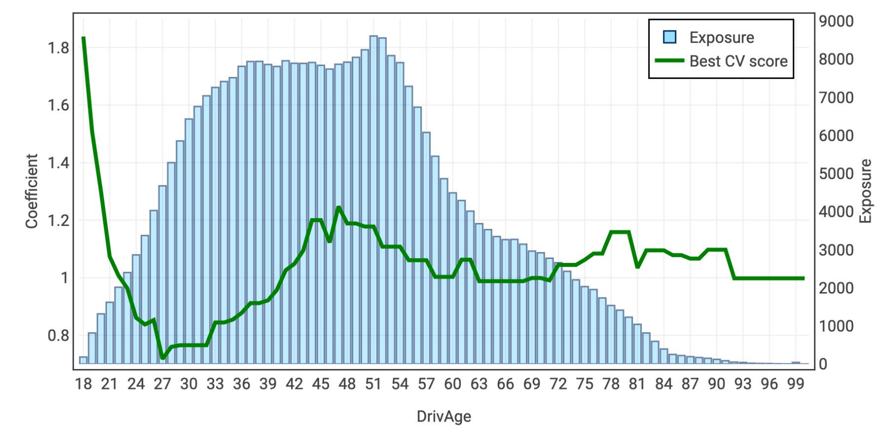
Performance is Not Enough

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

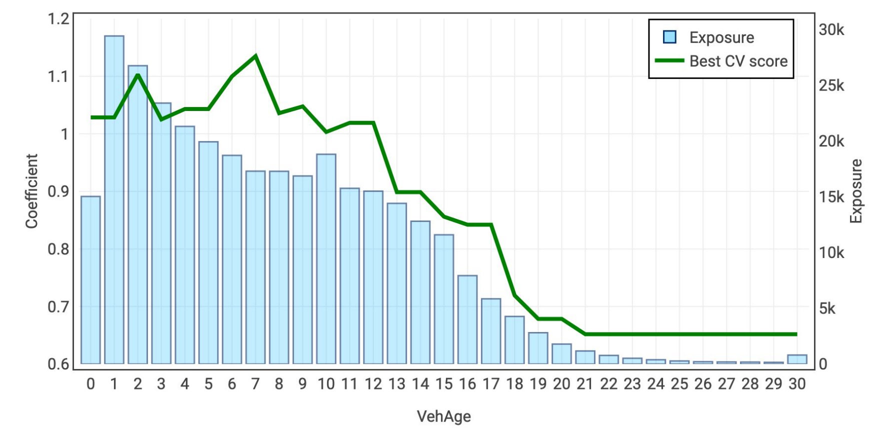
MGCV GAM would select **fewer splines.**

Derivative lasso selects a **higher penalty**





(a)

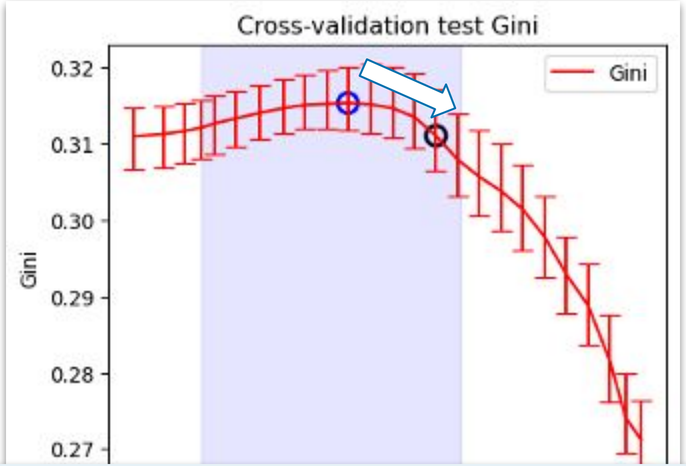


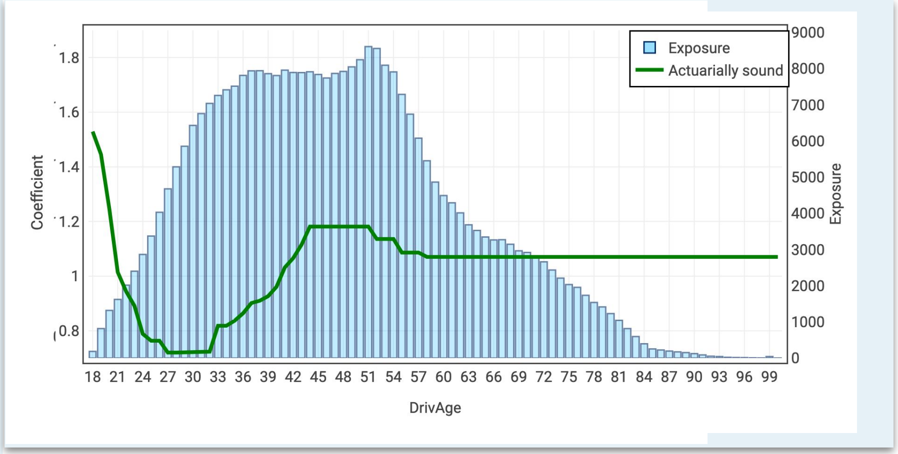
(b)

Applying Judgment

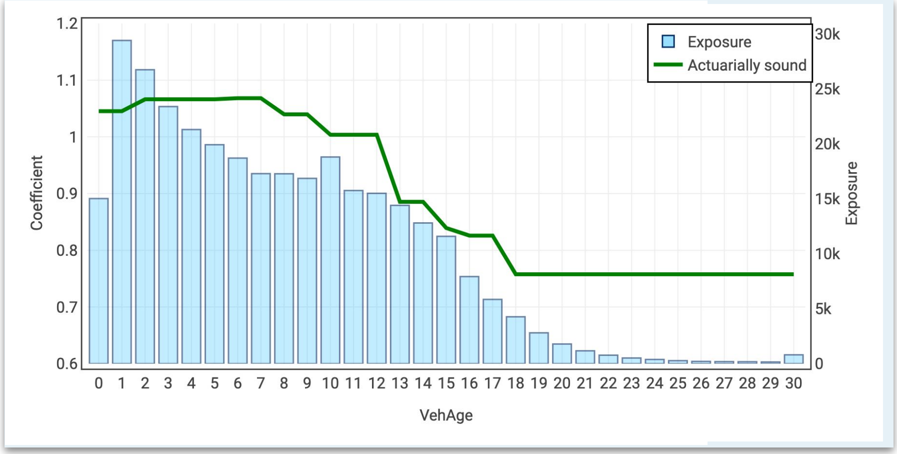
Actuaries should select a higher credibility threshold if this results in a more actuarially sound model.

This applies to all variables.





(a)

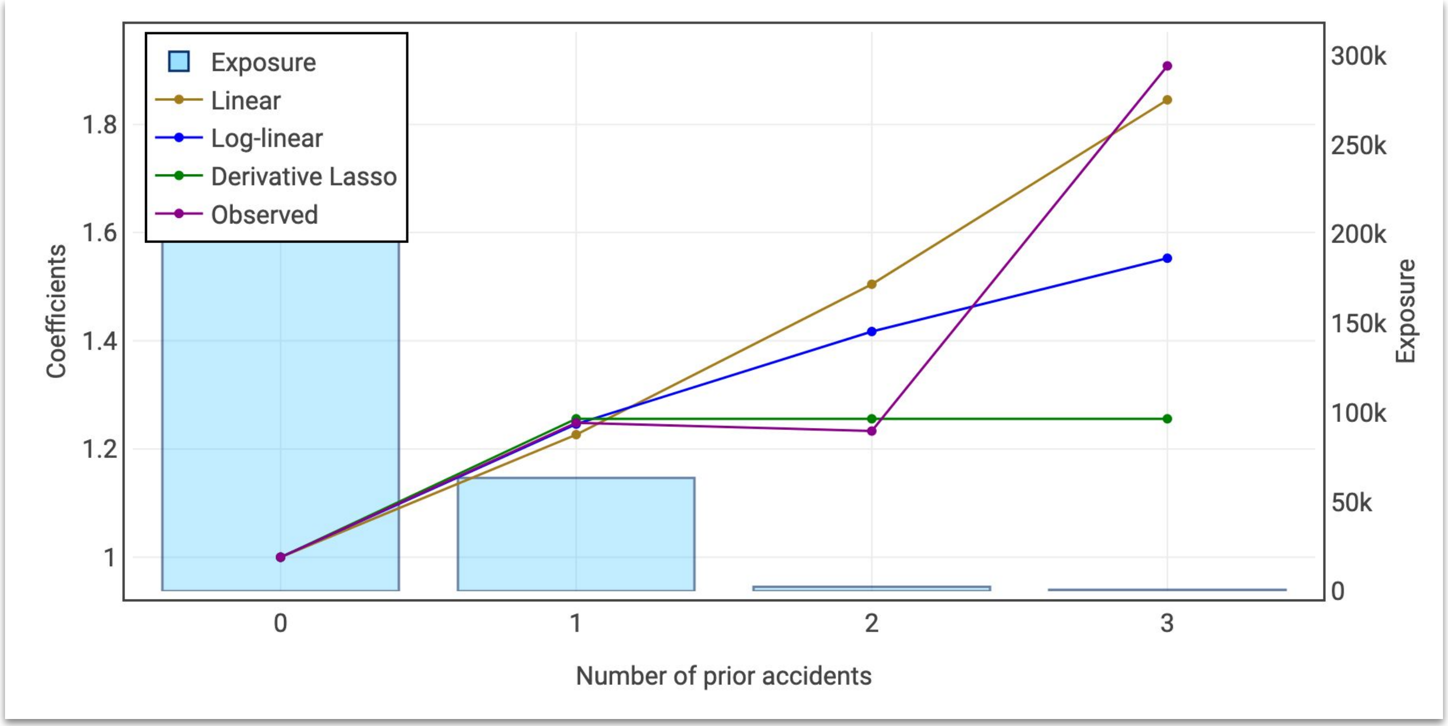


(b)

Derivative Lasso separates Statistical and Judgmental decisions

AKUR8

Feature Engineering is a Judgmental Modeling Choice



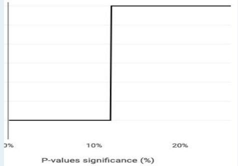
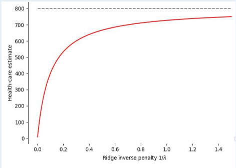
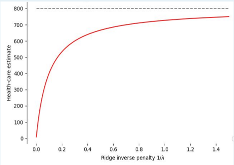
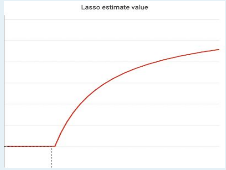
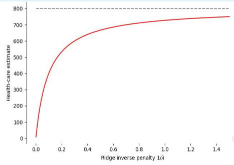
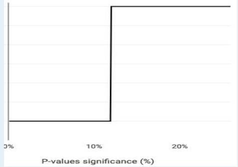
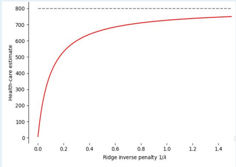
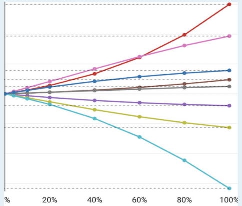
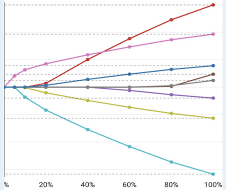
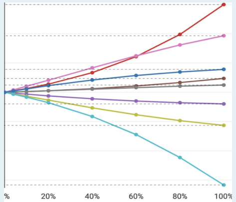
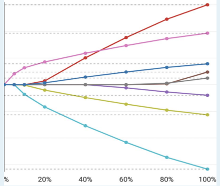
Separating Judgmental and Technical Decisions

- **Technical decisions:** modeling decisions that aim to improve the quality of the model's fit to the data.
- **Judgment-based decisions:** modeling decisions that aim to incorporate the modeler's opinions and experience, often based on judgment or business criteria.

These two decisions overlap in GLM, Ridge, and Lasso models.

The **distinction** between technical and judgment-based decisions is often **ambiguous in other types of penalized GLMs**, but it becomes **well-defined in derivative lasso**.

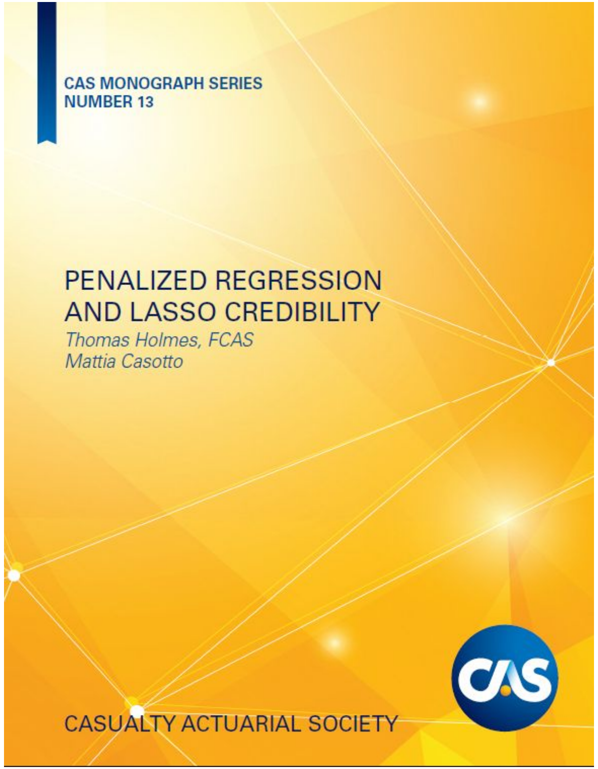
A Broader Comparison

<table>
<tr><th></th><th>GLM Significance</th><th>Buhlmann Credibility</th><th>Ridge</th><th>Lasso</th><th>MGCV GAM</th><th>Derivative Lasso</th></tr>
<tr><td>Inclusion Threshold</td><td>Yes</td><td colspan="2">No</td><td>Yes</td><td>No</td><td>Yes</td></tr>
<tr><td>Credibility Weighting</td><td>No</td><td colspan="5">Yes</td></tr>
<tr><td>Multivariate</td><td>Yes</td><td>No</td><td colspan="4">Yes</td></tr>
<tr><td>Judgmental Engineering</td><td colspan="4">Yes</td><td>Partial</td><td>No</td></tr>
<tr><td>Credibility Assignment</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Coefficient Path</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

Lasso Credibility extends Derivative Lasso

AKUR8

CAS Monograph 13: Penalized Regression and Lasso Credibility



Peer Reviewed CAS Monograph

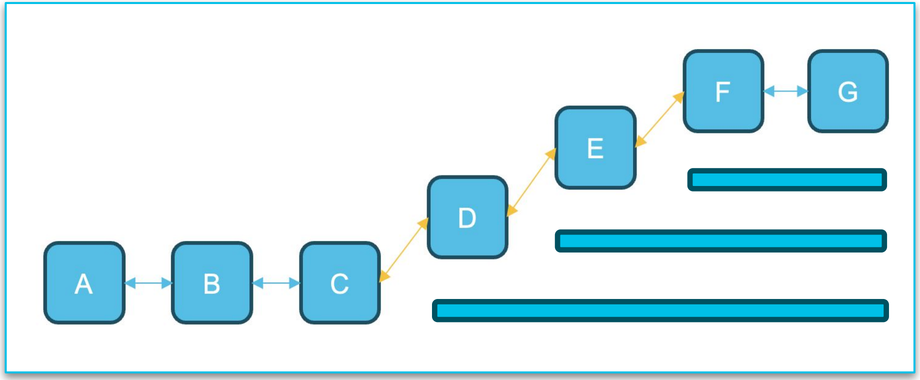
Includes Case Study

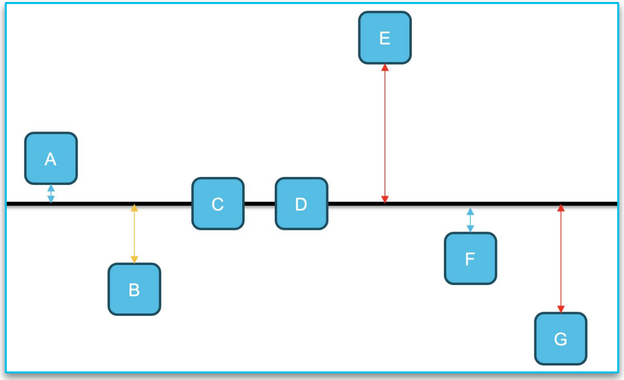
Resource for continuing education

Additional material and clarifications to be added “soon”

Aligning Penalization and Credibility

To intuitively apply penalization as credibility, all of our coefficients must represent categorical magnitude - **not slopes**.

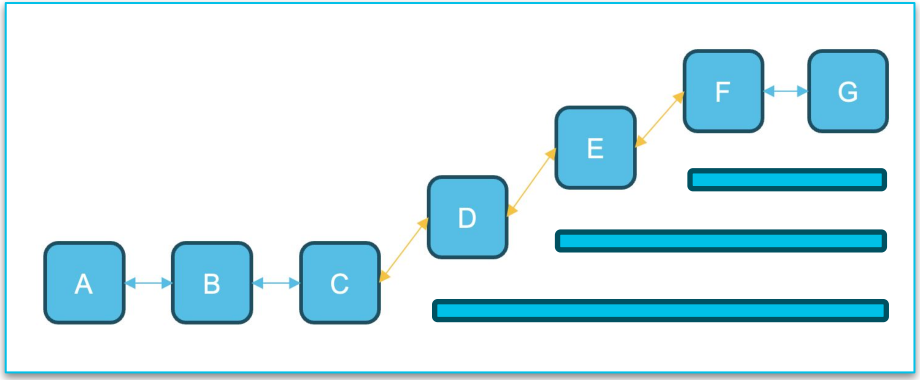


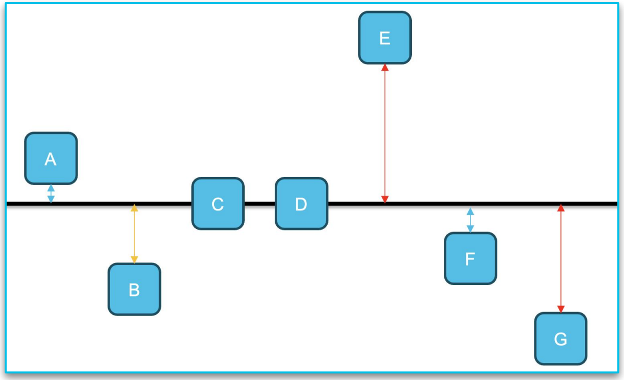


In the above examples, each coefficient could be evaluated individually through a credibility procedure (classical, Buhlmann).

A Stepwise Basis Function **Natively** Creates this Relationship

The basis function creates sufficiently granular breakpoints such that no manual groupings or manual engineering is performed.





“Normal” Lasso can be parameterized to reach a materially similar setup

The Offset as a Complement of Credibility

Decomposition of the Offset is essential to understand Lasso Credibility

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

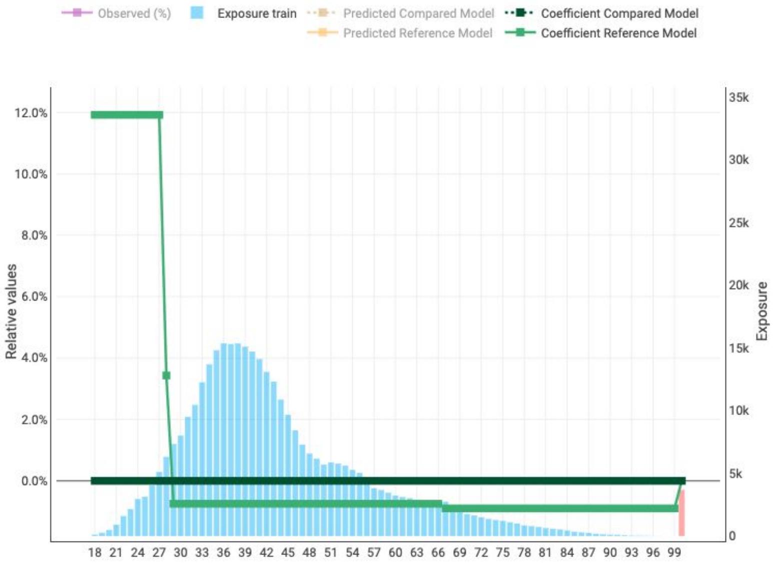
Complement

Category Definition

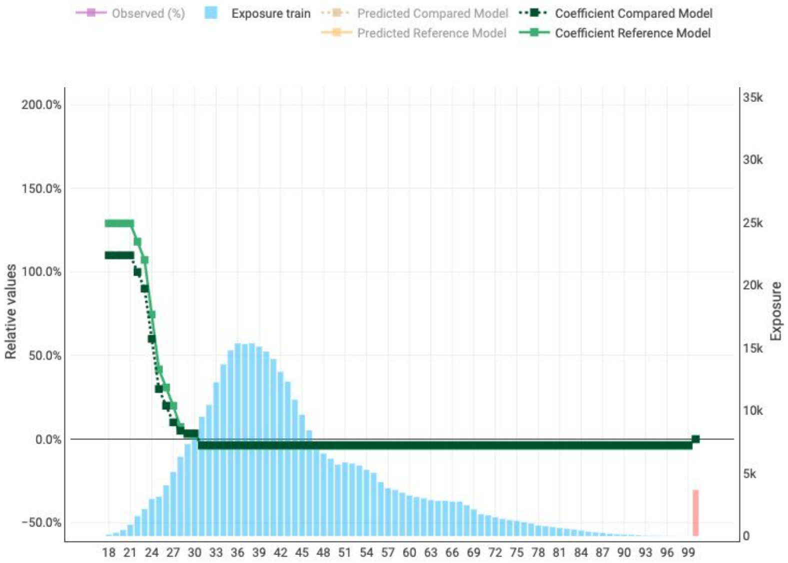
Modeled Coefficient

Visualization of Offset Decomposition

Without Visualizing Offset



With Visualizing Offset



The Offset as a Complement of Credibility

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

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

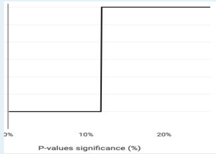
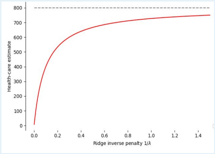
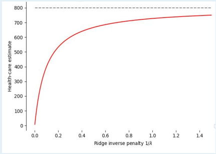
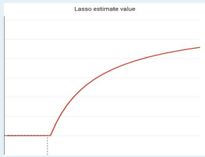
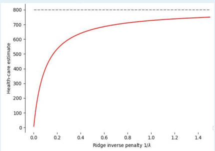
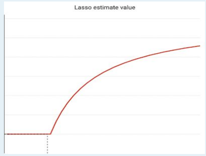
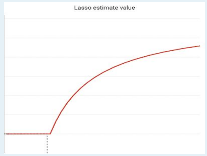
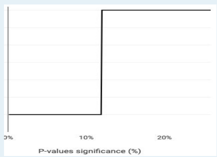
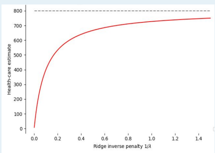
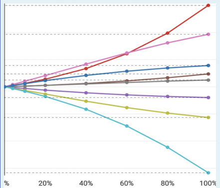
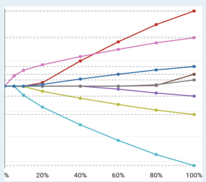
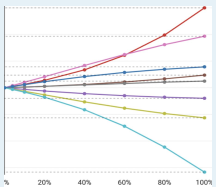
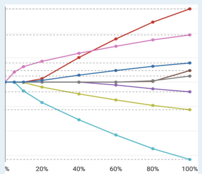
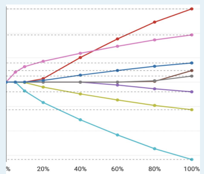
$$\text{Fully penalized coefficient} = \beta_{n,\text{offset}} + \beta_n = \beta_{n,\text{offset}} + 0 = \beta_{n,\text{offset}}$$

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

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

One Likelihood-based Credibility Standard

A Final Comparison

<table>
<tr><th></th><th>GLM Significance</th><th>Buhlmann Credibility</th><th>Ridge</th><th>Lasso</th><th>MGCV GAM</th><th>Derivative Lasso</th><th>Lasso Credibility</th></tr>
<tr><td>Inclusion Threshold</td><td>Yes</td><td colspan="2">No</td><td>Yes</td><td>No</td><td colspan="2">Yes</td></tr>
<tr><td>“Credibility” Weighting</td><td>No</td><td colspan="6">Yes</td></tr>
<tr><td>Multivariate</td><td>Yes</td><td>No</td><td colspan="5">Yes</td></tr>
<tr><td>Judgmental Engineering</td><td colspan="4">Yes</td><td>Partial</td><td colspan="2">No</td></tr>
<tr><td>Actuarial Credibility Procedure by Definition</td><td colspan="5">No</td><td>Partial</td><td>Yes</td></tr>
<tr><td>Credibility Assignment</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Coefficient Path</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

AKUR8 ACADEMY

Upcoming and Recent Replay Akur8 Academy Webinars

April 2, 2025 | 12PM ET - 6PM CET

[Integrating AI into Actuarial Work: Preserving Control and Applying Judgment]

Tom Holmes, Chief Actuary for the U.S. at Akur8

April 23, 2025 | 11AM ET - 5PM CET

[Gaining a Modeling Mindset]

Josh Meyers, Actuarial Data Scientist at Akur8

May 21, 2025 | 11AM ET - 5PM CET

[Drivers of Change: Uncover the Key Factors Behind Change in Ultimates]

Michael Henk, Actuary at Akur8
Bethany Cass, Consulting Partnerships Director at Akur8