Non-GLM Model Documentation

April 23, 2024

NAIC NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

Non-GLM Model Documentation

- Introduce CASTF White Paper and Appendices
- Key items from GAM Appendix
- Key items from Tree Based Model Appendix
- Future Other Penalized Regression Appendix
- Conclusion

NAIC Presenters

Sam Kloese, ACAS, MAAA, CSPA, CPCU
P&C Rate Modeling Actuary and Data Scientist
skloese@naic.org

April Yu, ACAS, MAAA
P&C Rate Modeling Actuary and Data Scientist
ayu@naic.org

Roberto Perez, FCAS, MAAA, CSPA
P&C Rate Modeling Actuary and Data Scientist
rperez1@naic.org

The Regulatory Review of Predictive Models

- [Regulatory Review of Predictive Models]
- White paper timeline
  - Adopted by CASTF 9/15/2020
  - Adopted by C Committee 12/8/2020
  - Adopted by EX Committee 4/14/21
- Introduction
  - *“Hopefully, this white paper helps bring more consistency to the art of reviewing predictive models within a rate filing and make the review process more efficient.”*
  - *“…this document is intended as guidance for state insurance regulators as they review predictive models.*
  - *“Nothing in this document is intended to, or could, change the applicable legal and regulatory standards…”*

White Paper Appendix B

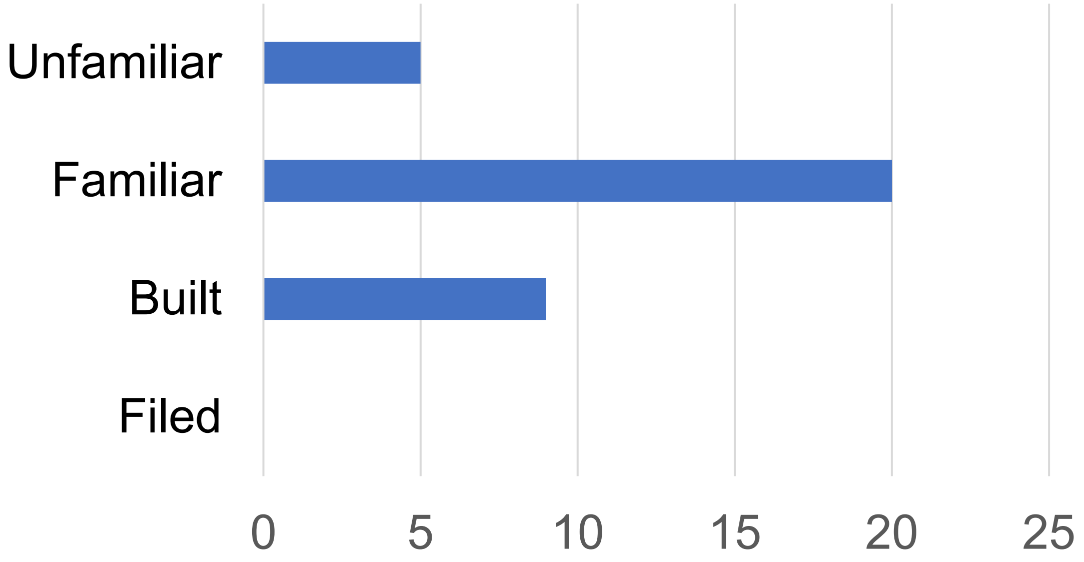
- *“This appendix identifies the information a state insurance regulator may need to review a predictive model used by an insurer to support a personal automobile or home insurance rating plan.”*
- Includes a list of information elements useful for a model reviewer
- Table also provides “Level of Importance to the Regulator’s Review”
  - Level 1: Necessary to begin the review
  - Level 2: Necessary to continue the review (with the exception of basic models)
  - Level 3: Necessary where concerns have been raised
  - Level 4: Necessary when the information in Level 1 - Level 3 have not resolved concerns
- *“If the model is not a GLM, some listed items might not apply…”*

Additional White Paper Appendices

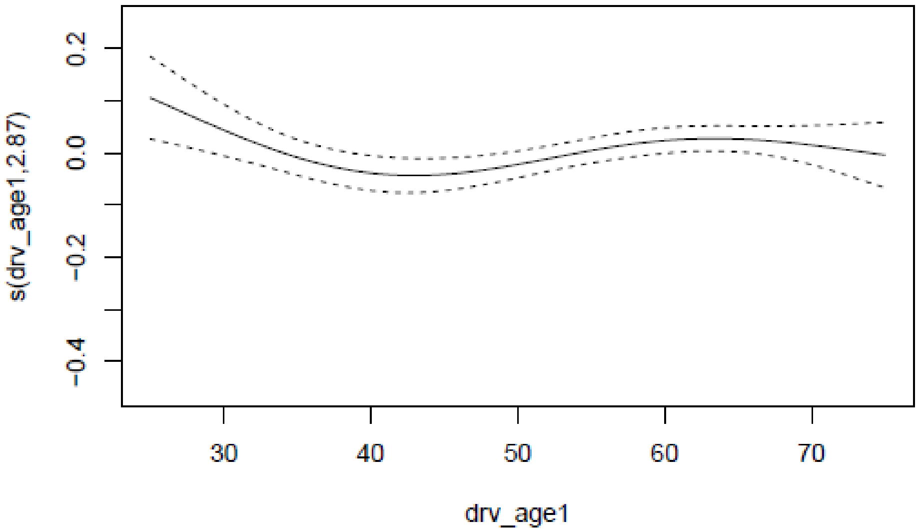
- Tree-based Models Appendix *(Adopted Summer 2022)*
  - Gradient Boosting Machines
  - Random Forest
- GAM White Paper Appendix *(Adopted Spring 2023)*
  - Generalized Additive Models including smoothed terms (mgcv package in R)
- Other Penalized Regression Appendix *(Not Yet Drafted)*
  - Elastic Net (Lasso, Ridge)
  - Accurate GLM (AGLM)
  - Derivative Lasso models

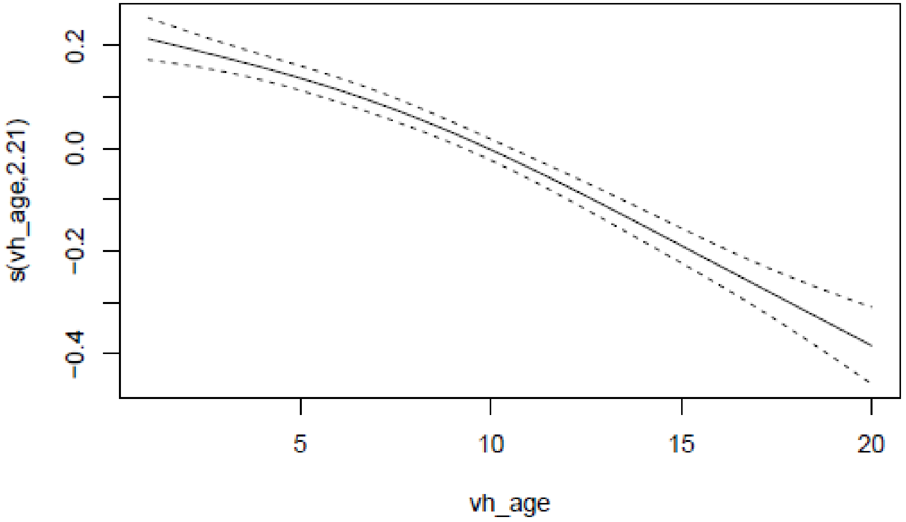
How familiar are you with GAMs built with mgcv in R?

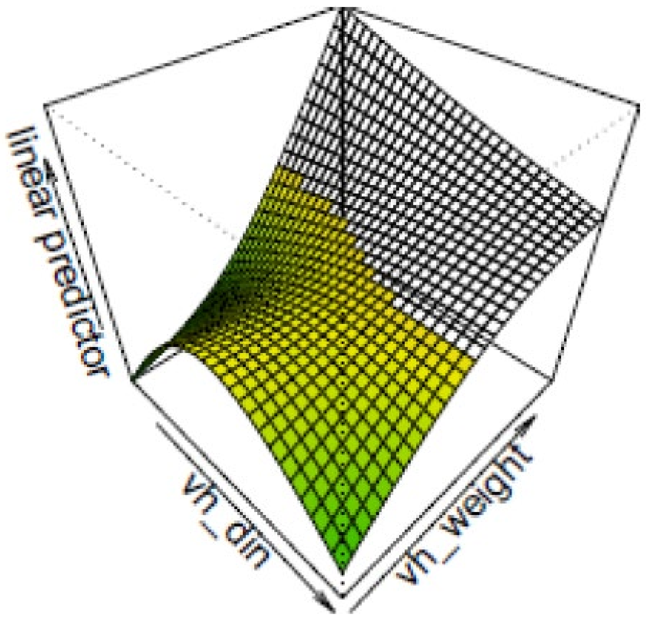
- I don’t know anything about GAMS.
- I am familiar, but I haven’t built one.
- I have built this type of model but never filed one.
- I have built this type of model and filed it for regulatory review.



Generalized Additive Models







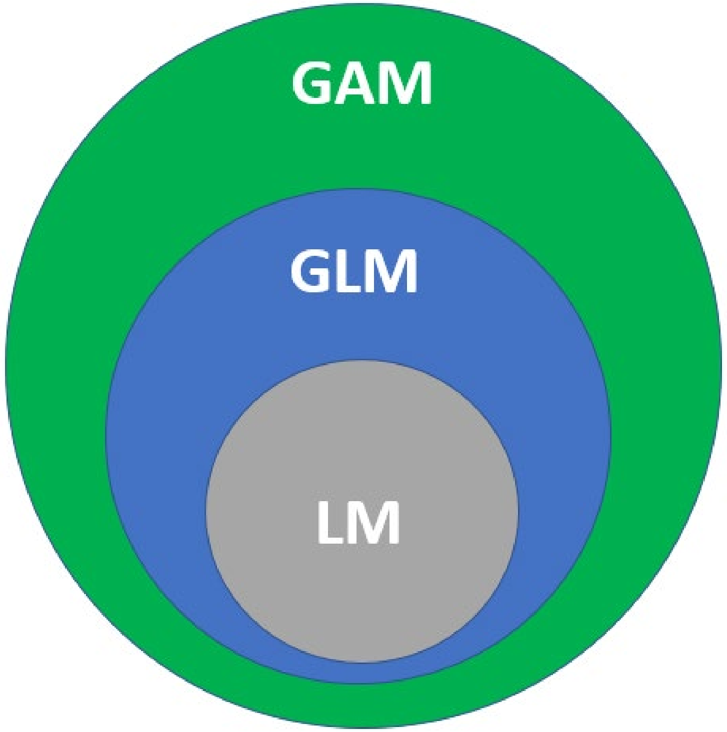
Similarity to GLMs

- GAMs are an extension of GLMs
- GAMs have many of the same elements
  - Multiple terms in the Regression functions to model the target variable
  - Allows selecting a distribution from the exponential distribution family (Poisson, Gamma, Tweedie, etc.)
  - Link Function defines the relationship between the linear predictor and the mean (log link, logistic link, etc.)
  - Offset terms can be added
  - Records can be weighted (exposures in frequency model, etc.)

```
gam_final <- gam(claim_count ~ pol_coverage + pol_usage +
                     s(drv_age1, k = 4) + s(vh_age, k = 4) +
                     te(vh_din, vh_weight, k = 3),
                   family = poisson(link = "log"),
                   offset = log(exposures),
                   data = training_data)
```

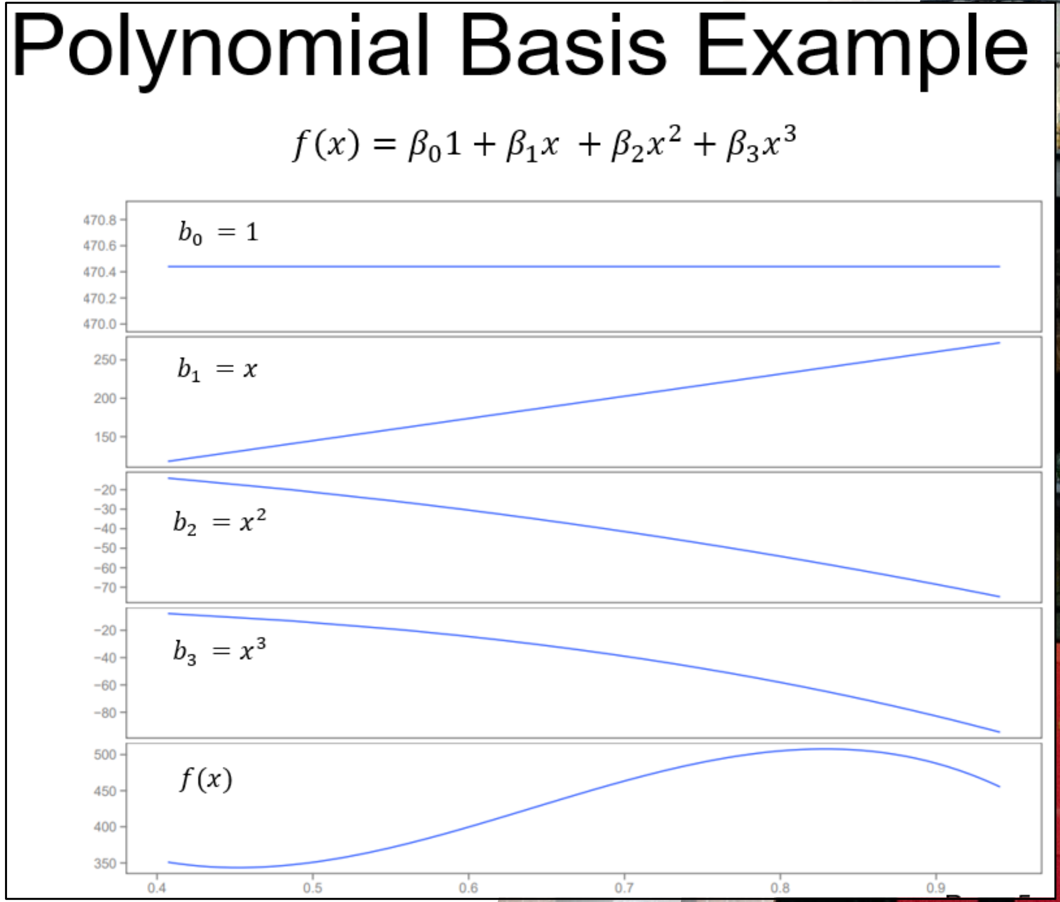
Similarity to GLMs

- GAM is like a GLM with the addition of smoothed terms
  - LM (Least squares): $\mu = \beta_0 + X_1\beta_1 + \ldots$
  - GLM: $g(\mu) = \beta_0 + X_1\beta_1 + \ldots$
  - GAM: $g(\mu) = \beta_0 + X_1\beta_1 + \ldots + f_1(X_1) + \ldots$
- LM to GLM to GAM
  - LM is a special case of GLM
    - Distribution: Normal
    - Link Function: Identity
  - GLM is a special case of GAM
    - No smoothed terms



Smooth Functions

- Smooth function are comprised of basis functions
- Modeling software allows you to set the type and number of the basis functions
- The overall impact of the smooth can be visualized and analyzed
- There are many types
  - Thin Plate
  - Cubic Splines
  - Random Effect
  - P Splines
  - Factor smooths

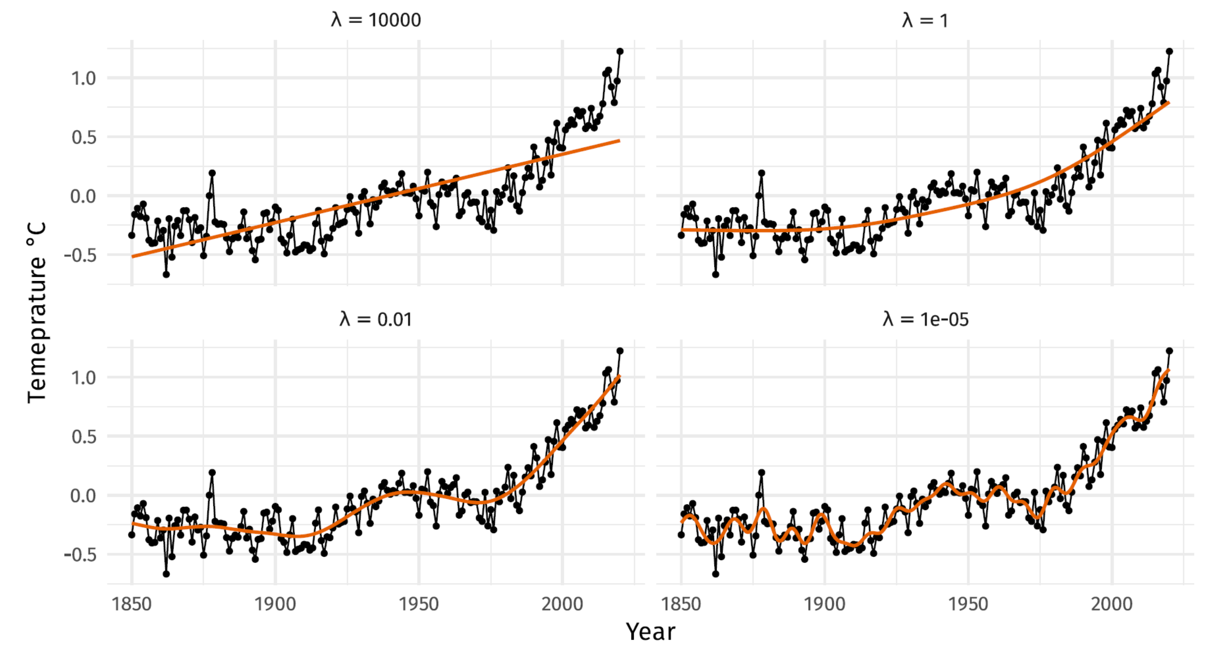


GAM is a type of Penalized Regression

- GAM Penalized Log-Likelihood
  - The smoothing parameter λ controls the penalty for the wiggliness of the model
  - The λ balances model fit vs. model complexity
    - High value: less wiggly
    - Low value: more wiggly, more responsive

$$L_p = \underbrace{L(\beta)}_{\text{Maximum Likelihood as in the GLM}} - \underbrace{\frac{1}{2}\lambda\beta^T S\beta}_{\text{Penalty to discourage overfitting - wiggliness}}$$

HadCRUT4 time series



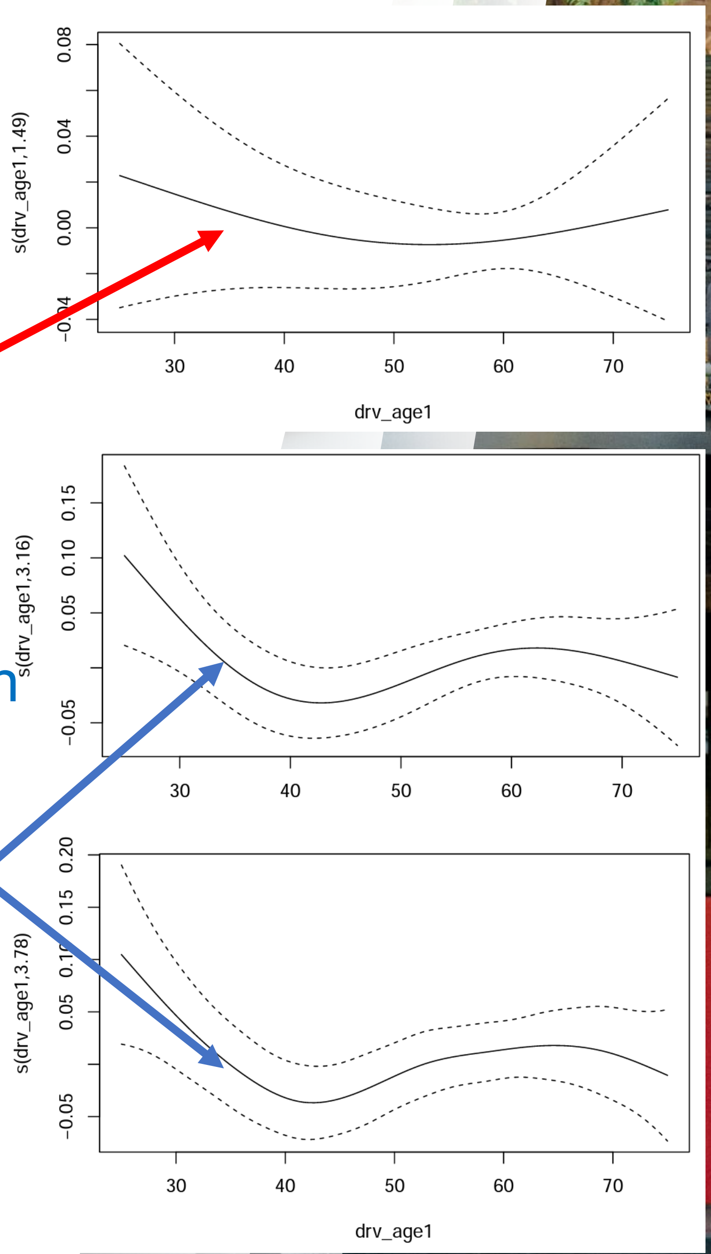
Smoothed Term Plots

- Focus on the reasonability of the aggregate smooth [Level 1 item]
  - Does the shape match the rational explanation?
- Place less focus on smooth type and underlying basis functions [Level 4 item]
- Consider if the confidence intervals are extremely wide
- Consider if the smooth seems overly noisy or overly smooth
- Consider if the smooth appears like it will extrapolate correctly
  - Look at the far left and far right sides
  - Look at areas with thinner data

Extremely wide confidence intervals

Fails horizontal line test

Different smooth types or more basis functions are not necessarily materially different



Smoothed Term Approximate P-values

- Approximate p-values are provided by the mgcv package in R
- Smoothed term p-values don't account for uncertainty in λ
- P-values are biased low, a lower threshold may be appropriate

```
##
## Family: poisson
## Link function: log
##
## Formula:
## claim_count ~ pol_coverage + pol_usage + s(drv_age1, k = 4) +
##     s(vh_age, k = 4) + te(vh_din, vh_weight, k = 3)
##
## Parametric coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -1.17696    0.18626  -6.319 2.63e-10 ***
## pol_coverageMedian1   -0.05899    0.03944  -1.496 0.134755
## pol_coverageMedian2   -0.13774    0.02885  -4.775 1.80e-06 ***
## pol_coverageMini      -0.59877    0.05396 -11.097  < 2e-16 ***
## pol_usageProfessional -0.40514    0.18800  -2.155 0.031163 *
## pol_usageRetired      -0.71978    0.18835  -3.822 0.000133 ***
## pol_usageWorkPrivate  -0.59133    0.18624  -3.175 0.001498 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                          edf Ref.df Chi.sq p-value
## s(drv_age1)            2.870  2.988  11.75 0.00653 **
## s(vh_age)              2.207  2.591 173.96 < 2e-16 ***
## te(vh_din,vh_weight)   6.453  7.073 176.90 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.0156   Deviance explained =  2.9%
## UBRE = -0.36299  Scale est. = 1         n = 79995
```

Concurvity Metrics

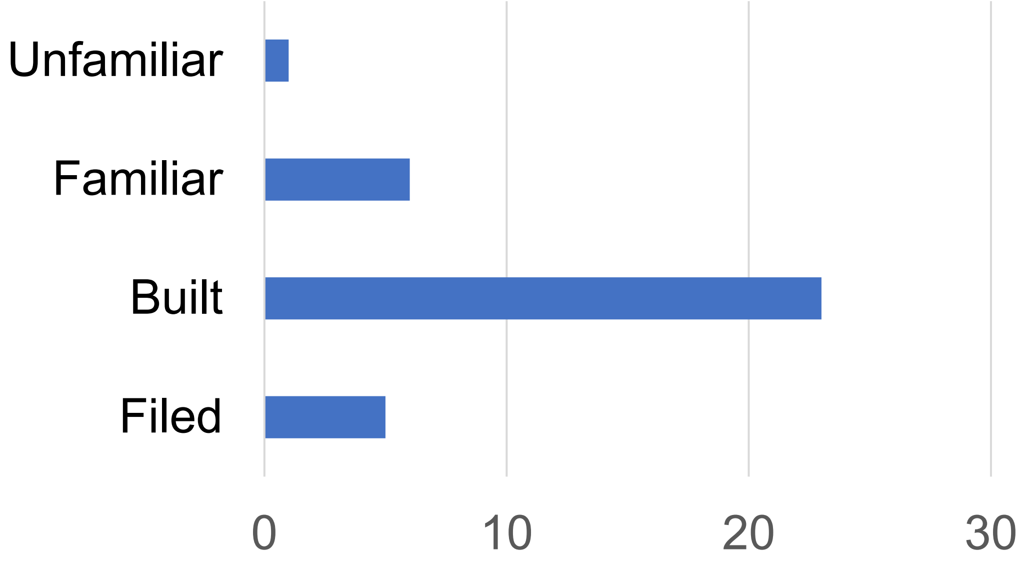
- Mgcv provides 3 versions of concurvity metrics: worst, observed, estimate
- Worst is the most pessimistic view
- Rule of thumb, a worst concurvity > 0.8 is too high for a smoothed term

```
concurvity(gam_final, full = TRUE)
```

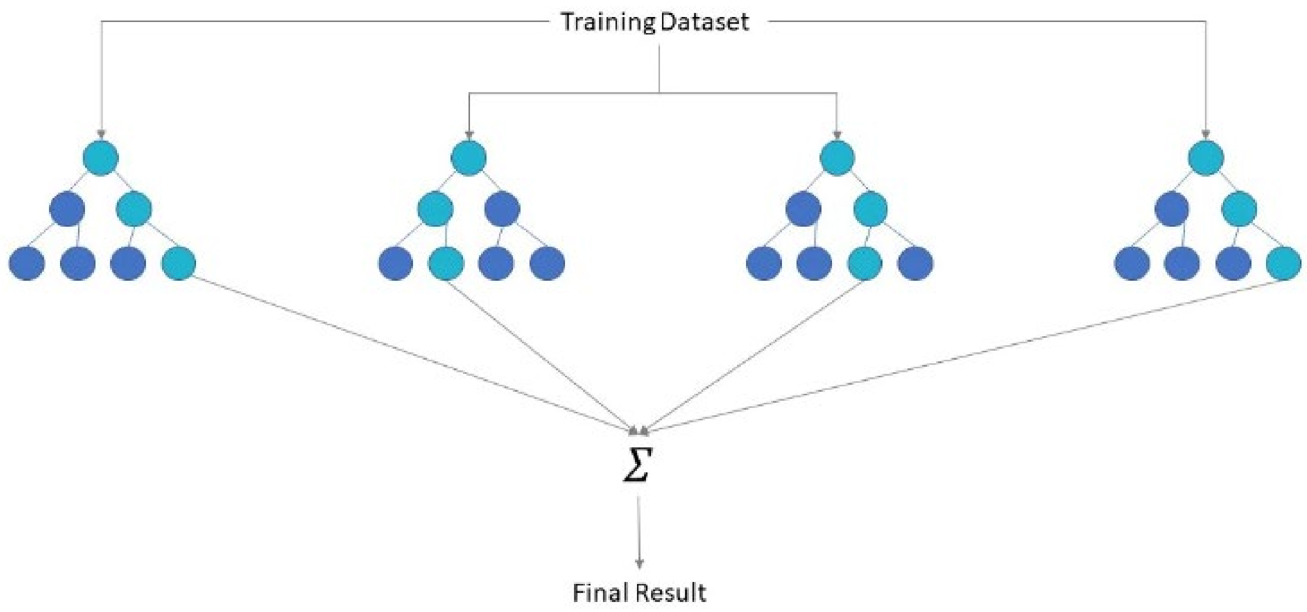
```
##               para s(drv_age1) s(vh_age) te(vh_din,vh_weight)
## worst    0.9990397  0.64082722 0.5583683            0.2826454
## observed 0.9990397  0.05038042 0.5504003            0.1831978
## estimate 0.9990397  0.42190878 0.5073782            0.1054095
```

How familiar are you with tree based models (Random Forest or GBM)?

- I don't know anything about RF or GBM.
- I am familiar, but I haven't built one.
- I have built this type of model but never filed one.
- I have built this type of model and filed it for regulatory review.



Tree Based Models

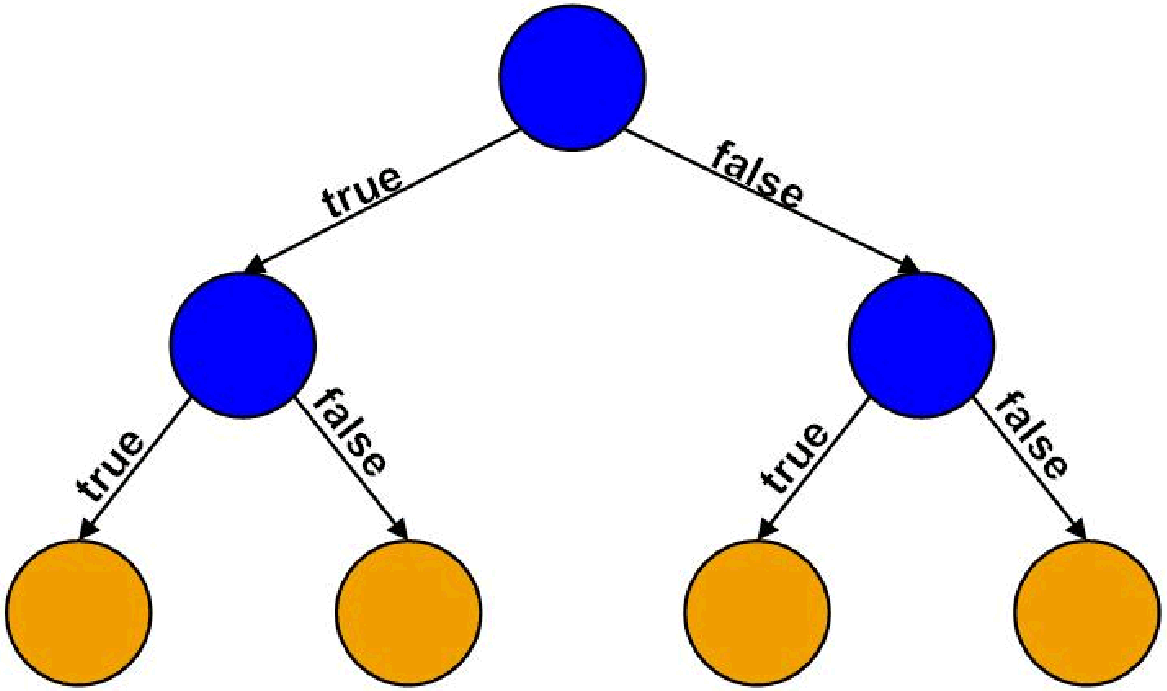


GBM

Random Forest

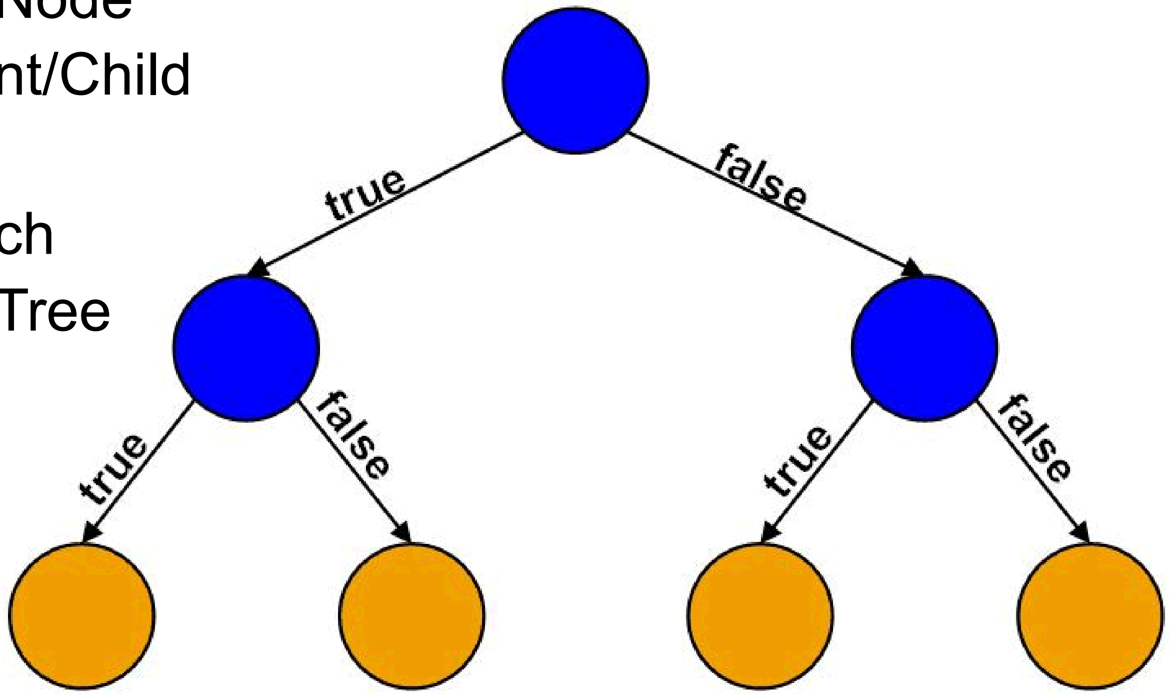
Single Decision Tree

- Easy to Understand
- Mimics how people make decisions
- Easily interpreted



Single Decision Tree

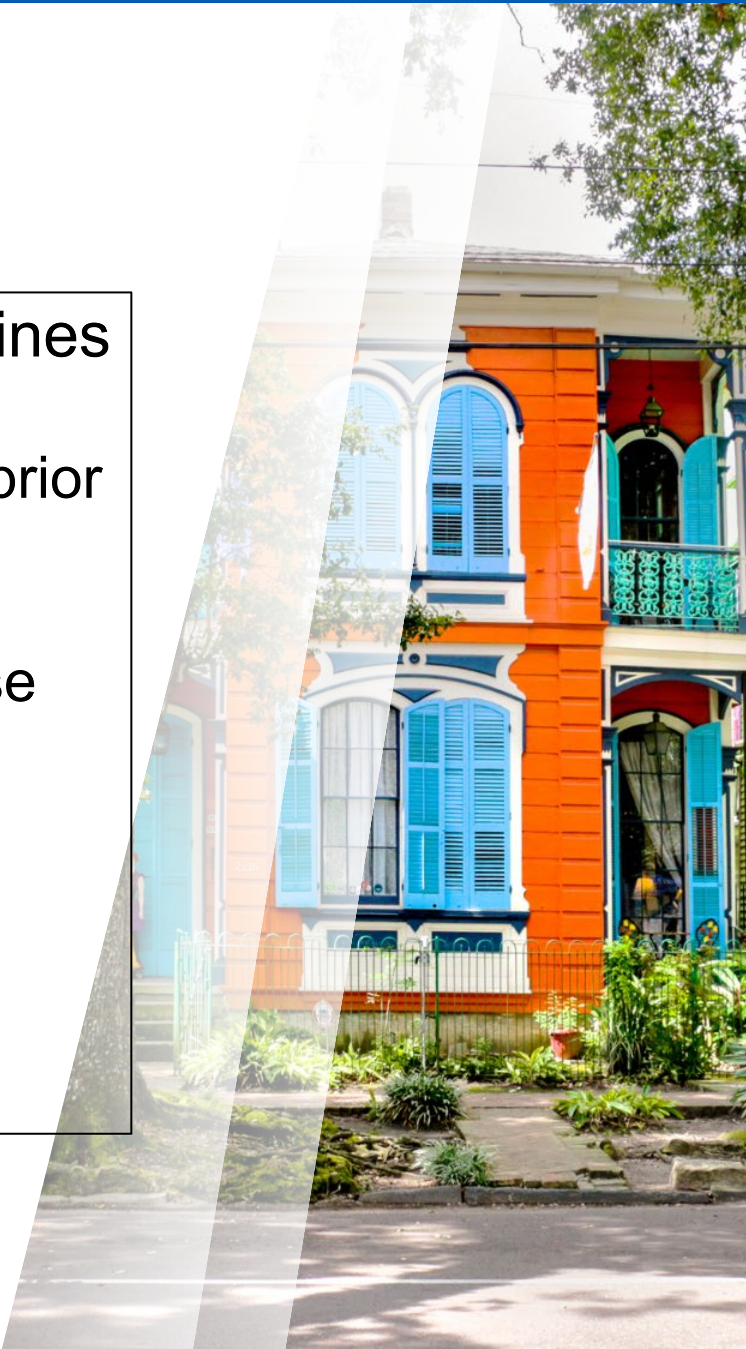
- Terminology
  - Nodes
    - Root
    - Sub-Node
    - Parent/Child
  - Splitting
    - Branch
    - Sub-Tree



Ensemble Tree Methods

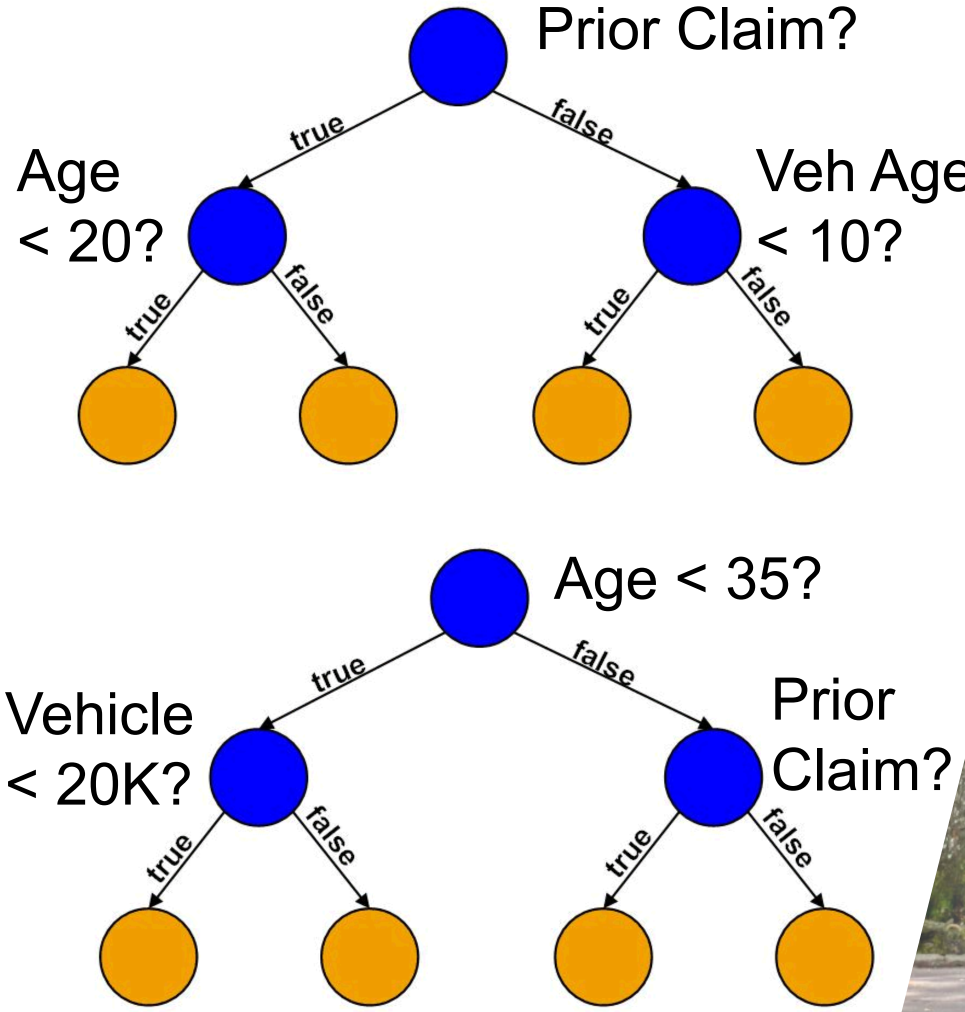
- Random Forest
  - Each Tree is based on a different bootstrap sample
  - Randomly chosen candidate variables at each split
  - Development of each tree is independent of the others
  - Final prediction is the average of the trees

- Gradient Boosting Machines
  - Subsequent trees are refined on errors from prior trees
  - Individual trees can be counterintuitive because they target residuals
  - Even more likely than Random Forest to be overfit



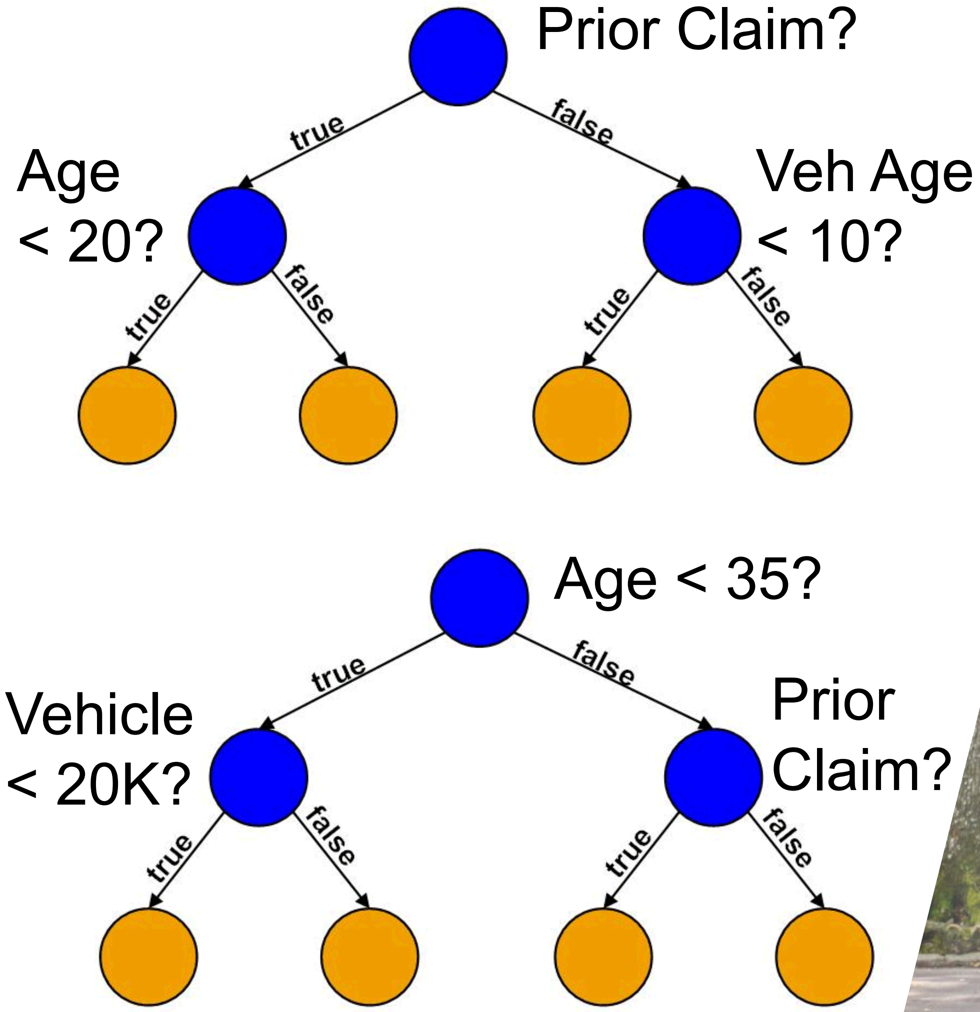
Ensemble Tree Challenges

- Interpretation gets difficult
  - Trees can get very deep (many splits)
  - There can be 100s or 1000s of trees
- Many GLM statistical tests no longer apply
- There are many hyperparameters
  - Selections may materially impact the model
  - Selections should be checked for reasonability



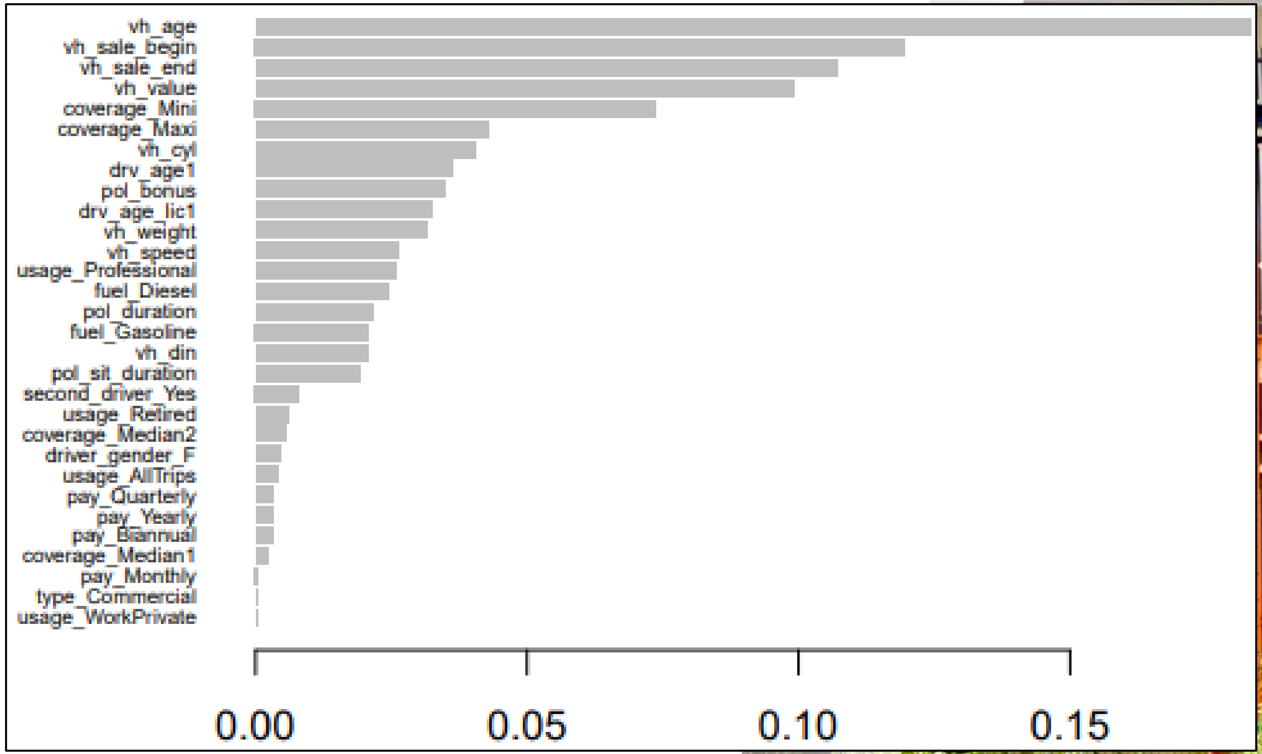
Ensemble Tree Hyperparameters

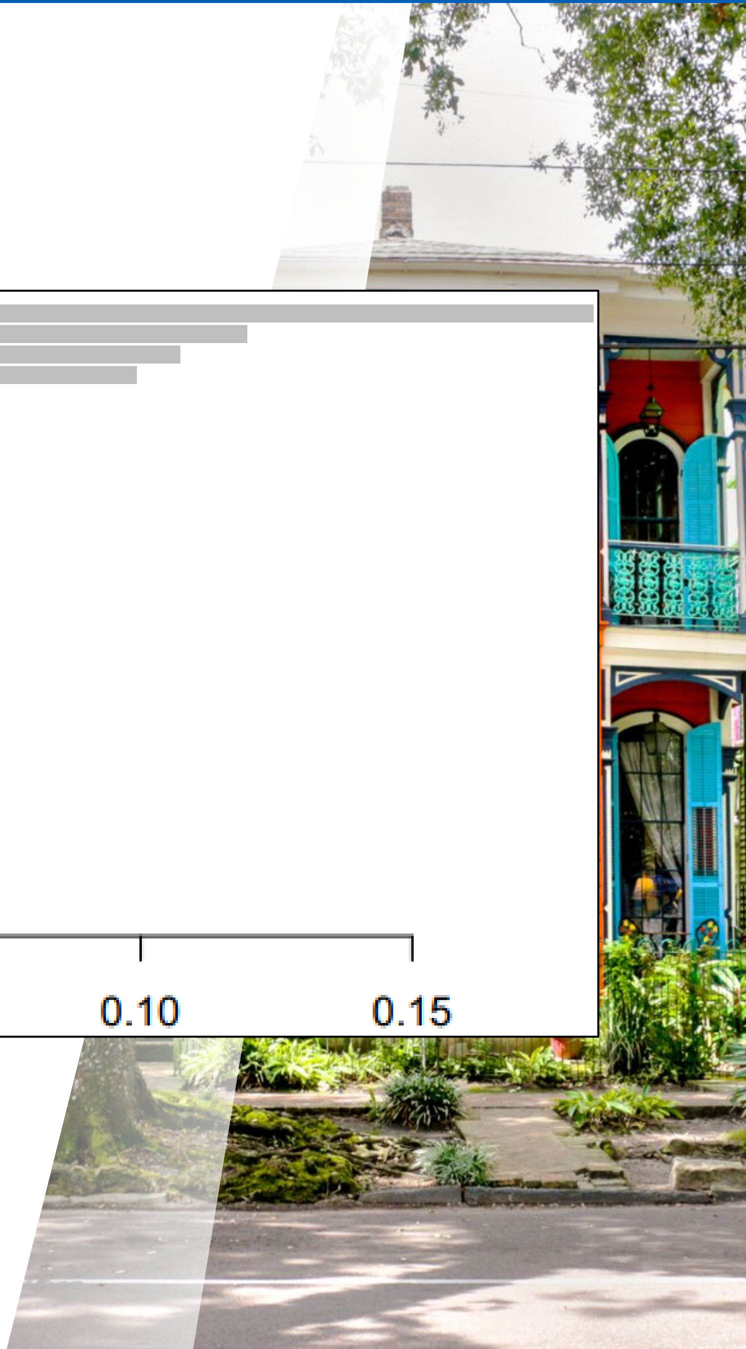
- Number of Trees
- Criteria on which to split
- Bootstrap sample size (% of rows)
- When to stop splitting
  - Max Tree Depth
  - Minimum Node Size
  - Max Leaf Nodes
- Random Variables for each split (# of columns)
- Learning rate (GBM only)



Variable Triaging

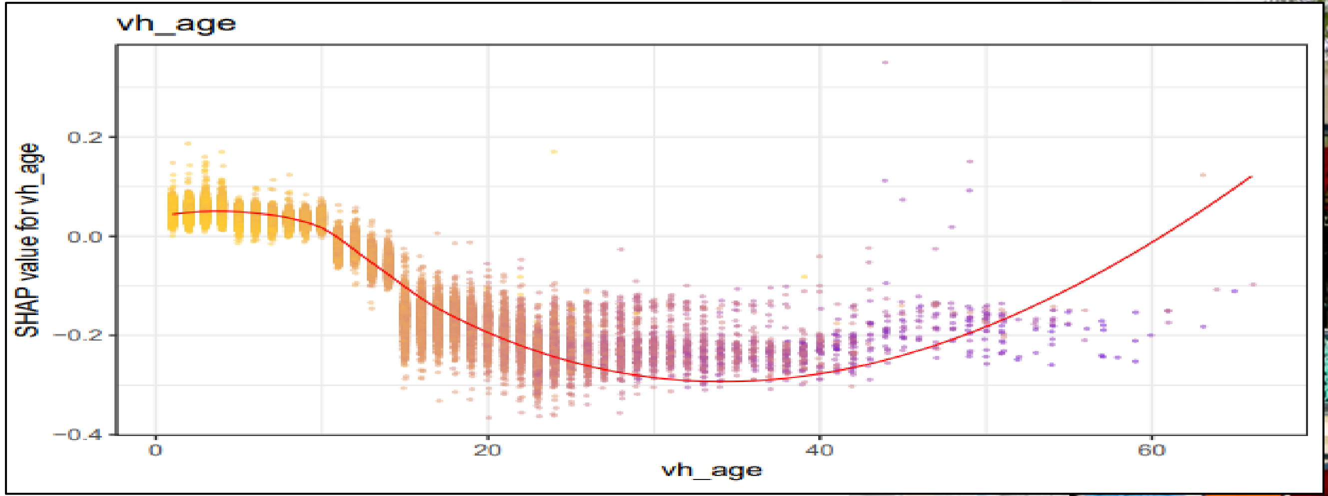
- Variable Importance Plots
  - Provide a measure of which variables are relatively more important than others
  - High importance variables should be evaluated as they will have the greatest impact on consumers
  - Low importance variables should be evaluated for whether there is a good reason to include them
    - Similar to questioning variables with high p-values in a GLM

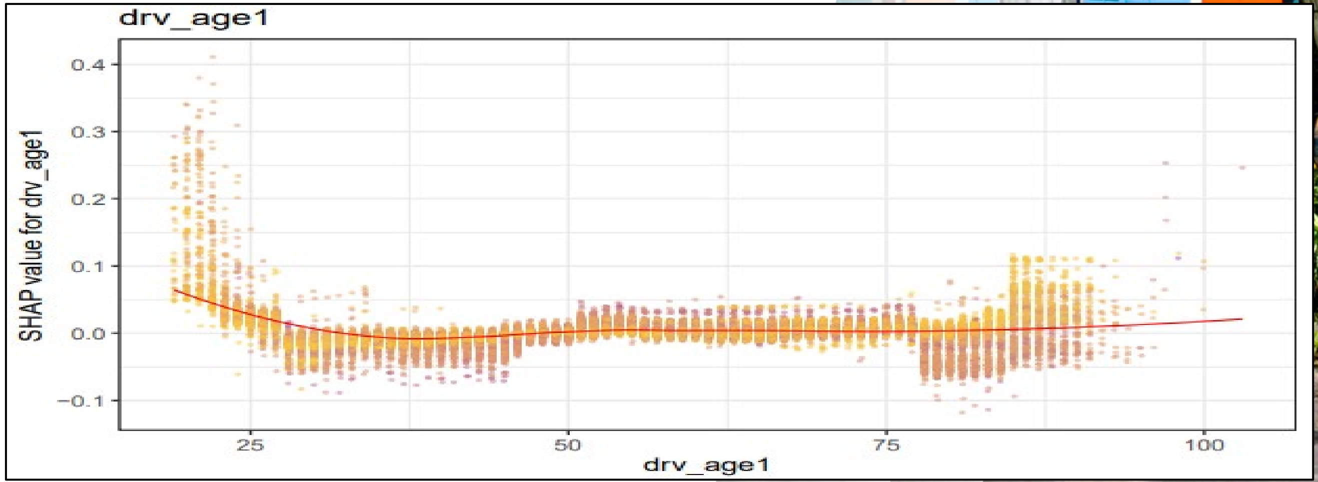




Interpretability Plots

- Assist in understanding the impact by variable
  - Partial Dependence Plots
  - Accumulated Local Effects plots
  - SHapley Additive exPlanations (SHAP)
- SHAP plots
  - How much that feature moves the prediction away from the overall average prediction.
  - >0, feature increases predicted value higher than average value
  - <0, feature decreases predicted value lower than average value





Assessing Overfit

- Review Hyperparameters
  - Number of trees should be large enough, but no larger
    - Look at plot to minimize OOB/Test Error or Deviance
  - Tree Complexity
    - Minimum node size should be set high enough for reasonable credibility
    - Rule of Thumb: Max depth of > 8 may be too high
  - Other hyperparameters should be disclosed and briefly commented on
    - Bootstrap sample size (% of rows)
    - Random Variables tried for each split (# of columns)
  - Criteria to split should match the model purpose (classification, regression)
- Review lift charts on test/holdout data

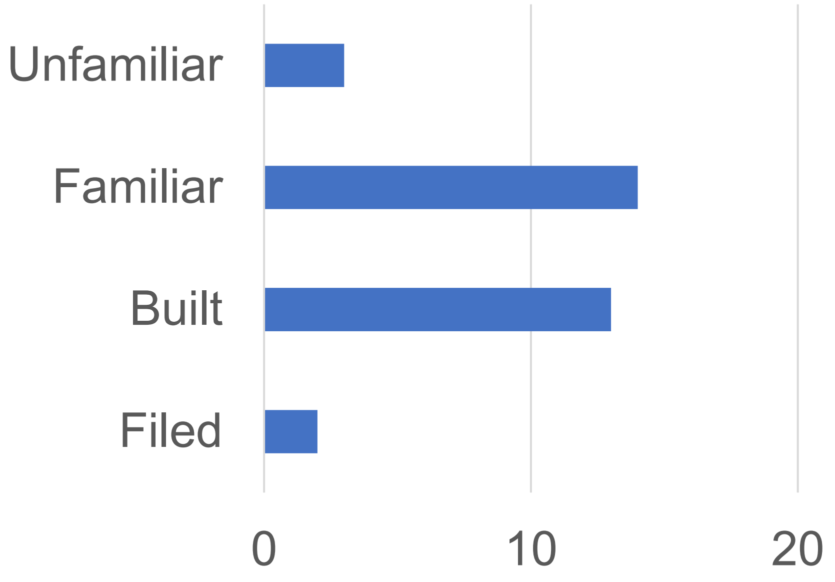
Auditability Challenges

- Tree Prediction Spot Check
  - Exhibits could be made for spot-checking against tree documentation
    - Input Predictors
    - Individual Tree Predictions
    - Overall Model Prediction
  - However, auditing every prediction for a book of business would still be extremely difficult.

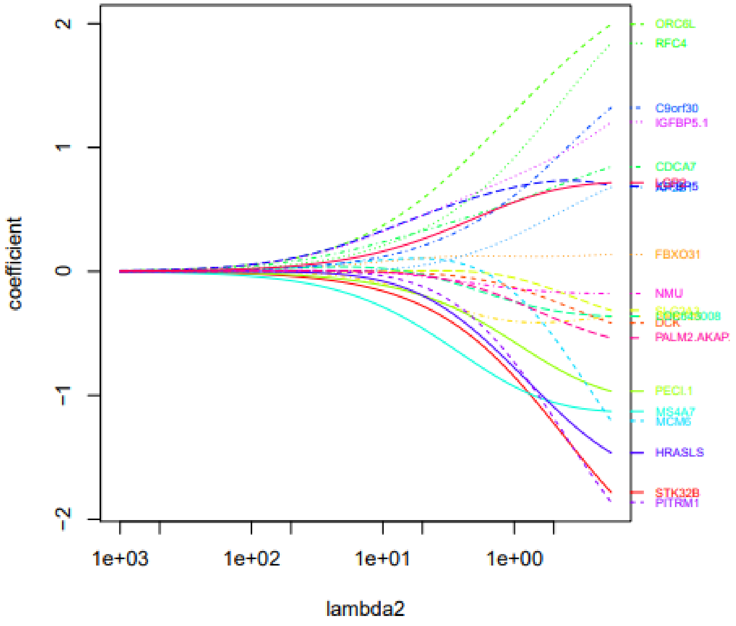
Sample Risk	Driver Age	Prior Claims	Vehicle Age	...	Tree 1	Tree 2	Tree 3	...	Model Prediction
1	16	0	5	...	$ 50.00	$ 40.00	$ 30.00	...	$ 40.00
2	17	0	6	...	$ 49.00	$ 39.20	$ 29.40	...	$ 39.20
3	18	0	2	...	$ 48.02	$ 38.42	$ 28.81	...	$ 38.42
4	19	1	3	...	$ 47.06	$ 37.65	$ 28.23	...	$ 37.65
5	20	0	9	...	$ 46.12	$ 36.90	$ 27.67	...	$ 36.90

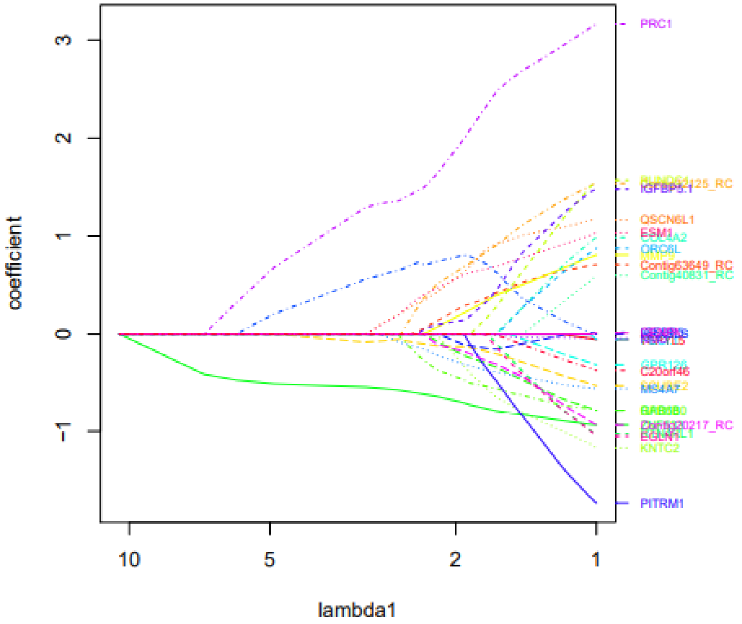
How familiar are you with penalized regression methods (elastic net, ridge, lasso, AGLM, etc.)?

- I don’t know anything about GAMS.
- I am familiar, but I haven’t built one.
- I have built this type of model but never filed one.
- I have built this type of model and filed it for regulatory review.



Other Penalized Regression Methods





Elastic Net **Ridge** **AGLM**

Lasso **Derivative Lasso**

Regular vs. Penalized Regression

J. R. Statist. Soc. B (2005)
67, *Part* 2, *pp.* 301–320

Regularization and variable selection via the elastic net

Hui Zou and Trevor Hastie

Stanford University, USA

[Received December 2003. Final revision September 2004]

Summary. We propose the elastic net, a new regularization and variable selection method. Real world data and a simulation study show that the elastic net often outperforms the lasso, while enjoying a similar sparsity of representation. In addition, the elastic net encourages a grouping effect, where strongly correlated predictors tend to be in or out of the model together. The elastic net is particularly useful when the number of predictors (p) is much bigger than the number of observations (n). By contrast, the lasso is not a very satisfactory variable selection method in the $p \gg n$ case. An algorithm called LARS-EN is proposed for computing elastic net regularization paths efficiently, much like algorithm LARS does for the lasso.

$$\hat{\beta} \equiv \underset{\beta}{\operatorname{argmin}}\left(\|y - X\beta\|^2 + \lambda_2\|\beta\|^2 + \lambda_1\|\beta\|_1\right)$$

Regular vs. Penalized Regression

J. R. Statist. Soc. B (2005)
67, *Part* 2, *pp.* 301–320

Regularization and variable selection via the elastic net

Hui Zou and Trevor Hastie

Stanford University, USA

[Received December 2003. Final revision September 2004]

Summary. We propose the elastic net, a new regularization and variable selection method. Real world data and a simulation study show that the elastic net often outperforms the lasso, while enjoying a similar sparsity of representation. In addition, the elastic net encourages a grouping effect, where strongly correlated predictors tend to be in or out of the model together. The elastic net is particularly useful when the number of predictors (p) is much bigger than the number of observations (n). By contrast, the lasso is not a very satisfactory variable selection method in the $p \gg n$ case. An algorithm called LARS-EN is proposed for computing elastic net regularization paths efficiently, much like algorithm LARS does for the lasso.

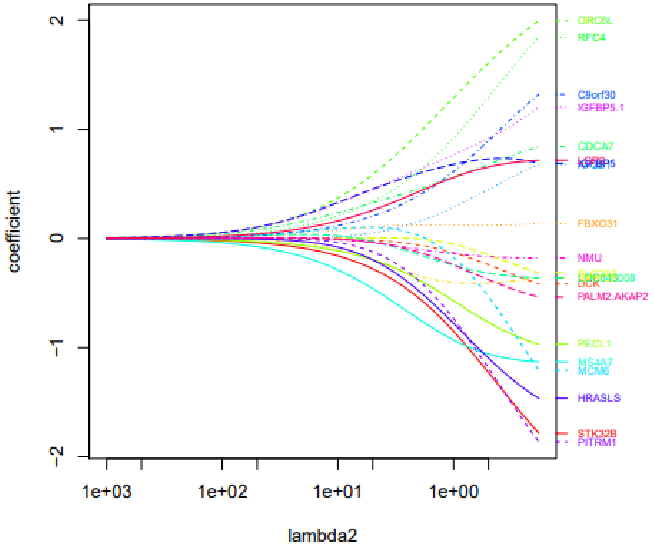
Keywords: Grouping effect; LARS algorithm; Lasso; Penalization; $p \gg n$ problem; Variable selection

Traditional Regression

$$\hat{\beta} \equiv \underset{\beta}{\operatorname{argmin}}\left(\boxed{\|y - X\beta\|^2} + \lambda_2 \|\beta\|^2 + \lambda_1 \|\beta\|_1\right)$$

Regular vs. Penalized Regression

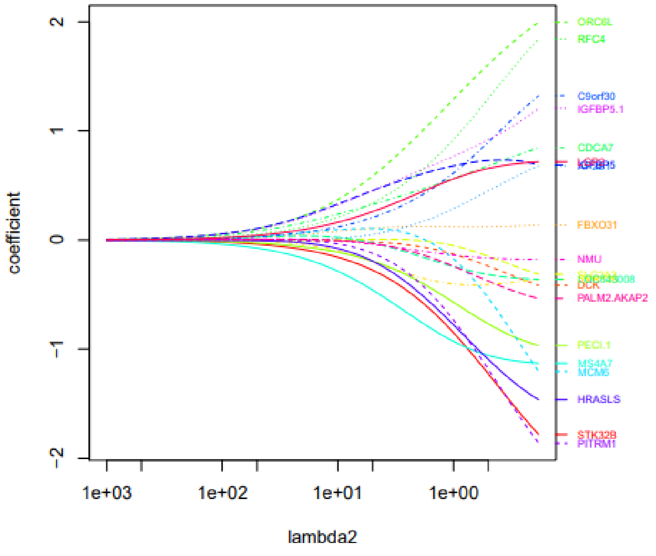
$$\hat{\beta} \equiv \underset{\beta}{\operatorname{argmin}}\left(\boxed{\|y - X\beta\|^2 + \lambda_2\|\beta\|^2} + \lambda_1\|\beta\|_1\right)$$



Regular vs. Penalized Regression (Ridge Regression)

$$\hat{\beta} \equiv \underset{\beta}{\operatorname{argmin}}\left(\boxed{\|y - X\beta\|^2 + \lambda_2 \|\beta\|^2} + \lambda_1 \|\beta\|_1\right)$$

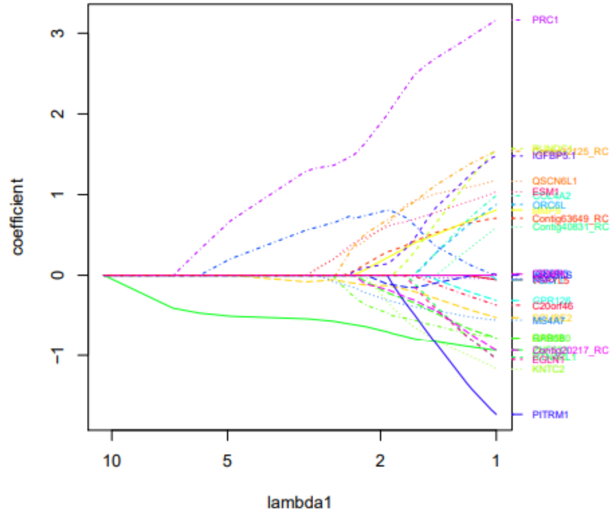
Ridge Regression



Regular vs. Penalized Regression (Lasso Regression)

$$\hat{\beta} \equiv \underset{\beta}{\operatorname{argmin}}\left(\boxed{\|y - X\beta\|^2} + \lambda_2 \|\beta\|^2 \boxed{+ \lambda_1 \|\beta\|_1}\right)$$

Lasso Regression



Regular vs. Penalized Regression (Elastic Net)

$$\hat{\beta} \equiv \underset{\beta}{\operatorname{argmin}}\left(\|y - X\beta\|^2 + \underbrace{\lambda_2}_{(1-\alpha)}\|\beta\|^2 + \underbrace{\lambda_1}_{\alpha}\|\beta\|_1\right)$$

- Elastic Net Regression can be thought of as a combination of Lasso and Ridge
  - α -> 0 → Closer to Lasso
  - α -> 1 → Closer to Ridge

Fitting an Elastic Net Model

- Alpha and Lambda Hyperparameters
- General Sample Process:
  - Define a range of hyperparameter values (i.e., alpha in [0,.2,.4,.6,.8.1] and lambda in [.0001,.001,.01,1]
  - Grid Search vs. Random Search
    - Grid search covers the full range but is more computationally intensive
  - Use Cross-Validation to optimize an objective function.

Regular vs. Penalized Regression (AGLM / Derivative Lasso)

- AGLM (R Package: aglm)
  - "a clear one-to-one relationship between the features and the response variable"
  - GLM + Regularization + Discretization + O/L variables
    - Discretization -> splitting numerical features into bins
    - O variables -> Reflects ordinal relationship between levels
    - L variables-> Ensures consistency between adjacent bins
- Akur8 GLMs
  - Derivative Lasso
  - Variations + Fitting Procedures

$$\text{Derivative Lasso}(\beta) = \sum_{j=1}^{p-1} |\beta_{j+1} - \beta_j|$$

Reviewing Penalized Regression Models

$$\hat{\beta} \equiv \underset{\beta}{\mathbf{argmin}}(||y - X\beta||^2 + \lambda_2||\beta||^2 + \lambda_1||\beta||_1)$$

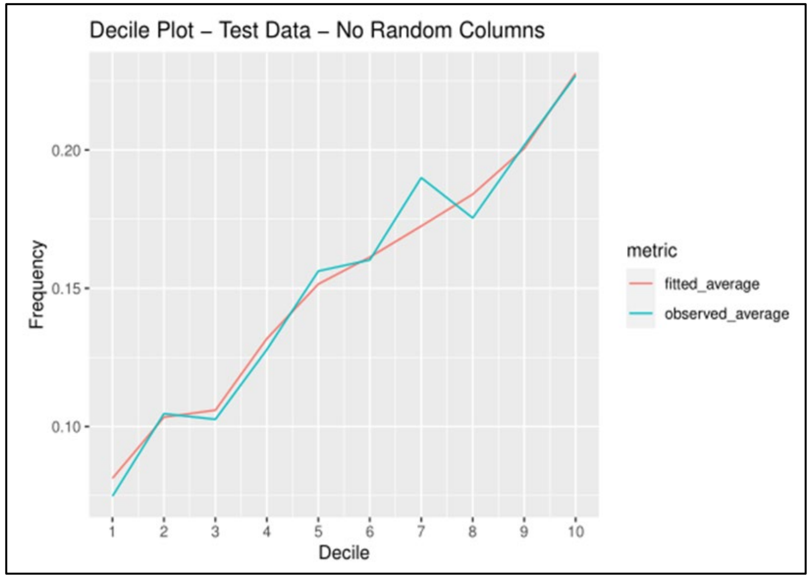
- Very similar to reviewing a GLM, however…
  - "standard errors are not very meaningful for strongly biased estimates such as arise from penalized estimation methods."
  - Penalized methods introduce bias when estimating coefficients, which becomes a major component of MSE
  - Confidence statement based on variance can be misleading

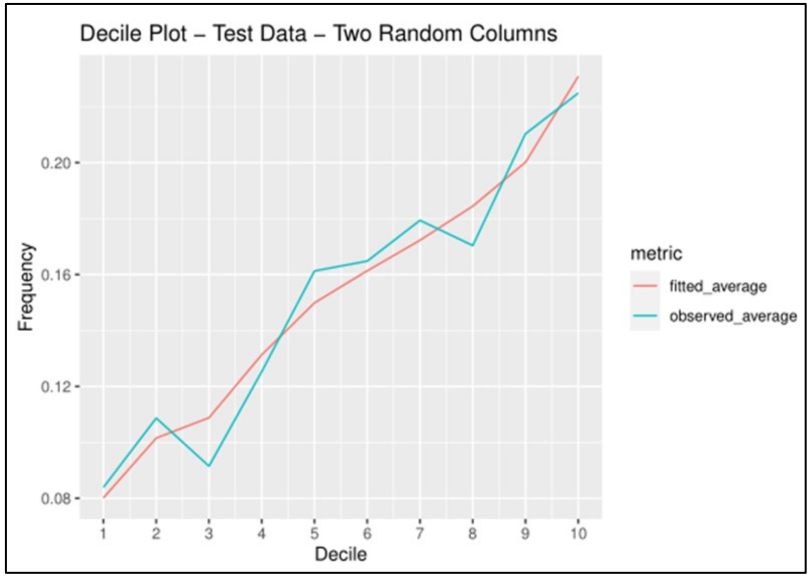
Final Model Lift Chart Is Not Enough

- Occasionally companies will reply that the overall model generalizes well to new data when asked about the significance of a specific variable
- A lift chart on holdout data may not look that bad if there is an insignificant variable included.

Final Model Lift Chart Is Not Enough

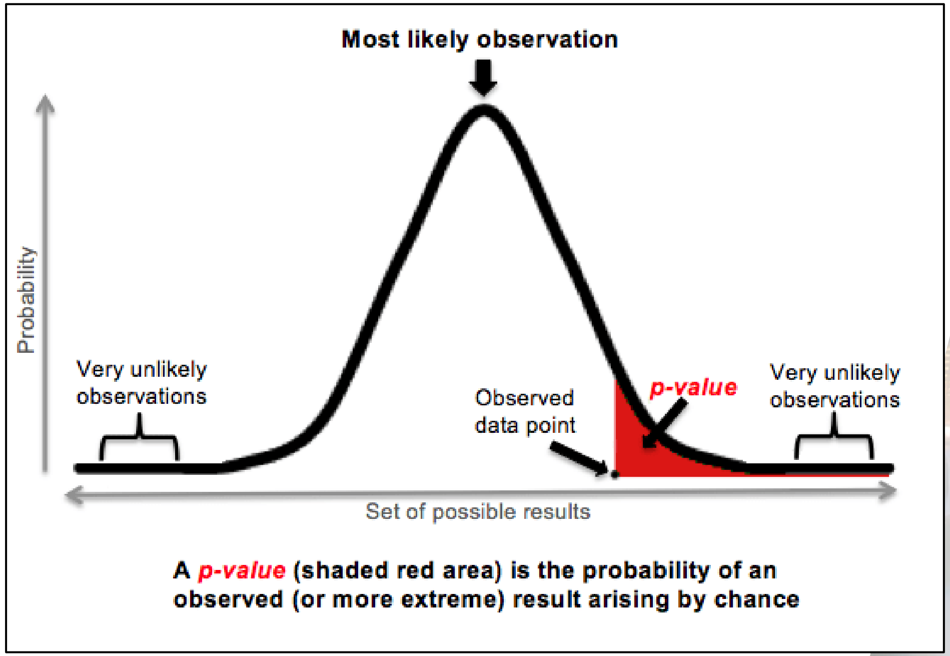
- Example
  - GLM was built, data included 100 columns with random #'s 1-5
  - 7 random # columns had P-value < 0.05
  - Model A was built excluding all random #s
  - Model B was built including 2 random # columns with lowest p-values
  - The decile plot for Model B doesn't look that bad!





P-Values

- P-Value
  - For a given statistical model when the null hypothesis is true, the P – value is the probability the model test statistic is equal to or more extreme than the actual observed results.



  - A p-value is NOT the probability that the null hypothesis is true

- For regression analysis, we test
  1.) H_0: $\beta_i = 0$
  2.) H_0: σ_i are equal

Reviewing Penalized Regression Models – p-values

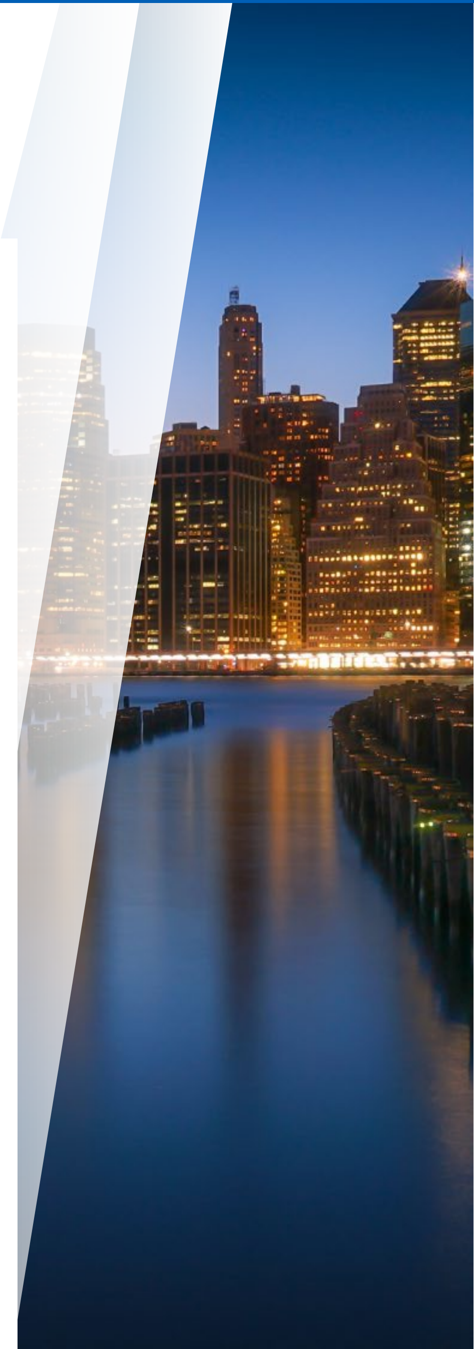
- R Packages
  - No p-values
    - Glmnet
    - HDM
    - BigLasso
    - lars
    - Caret
    - h2o
- Lassopv
  - Uses the regularization strength when each predictor enters the active set of regularization path for the first time as the statistic. (Only for Lasso)

P-value Alternatives

- P-values are a common metric for variable significance
- Other tests that may help address the question of significance
- Bootstrapping: Do variations to the data result in radically different coefficients?
- Cross Fold Validation: Are the coefficients consistent across folds?
- GLM Reference Model: What are the p-values from a similar GLM?

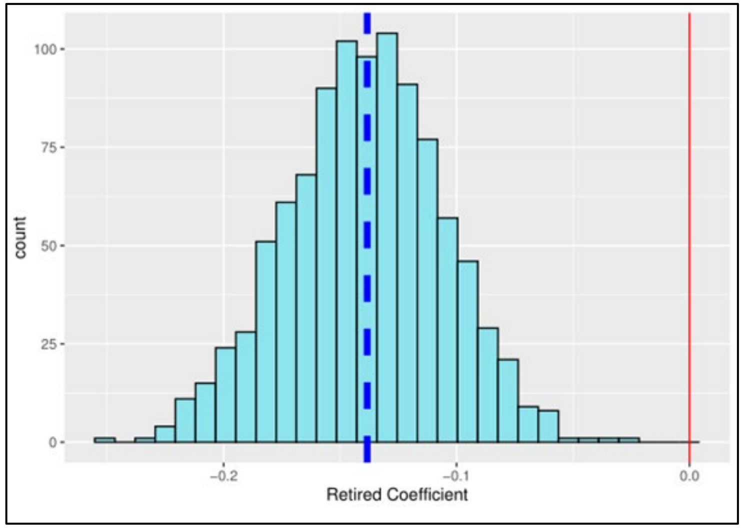
Bootstrapping

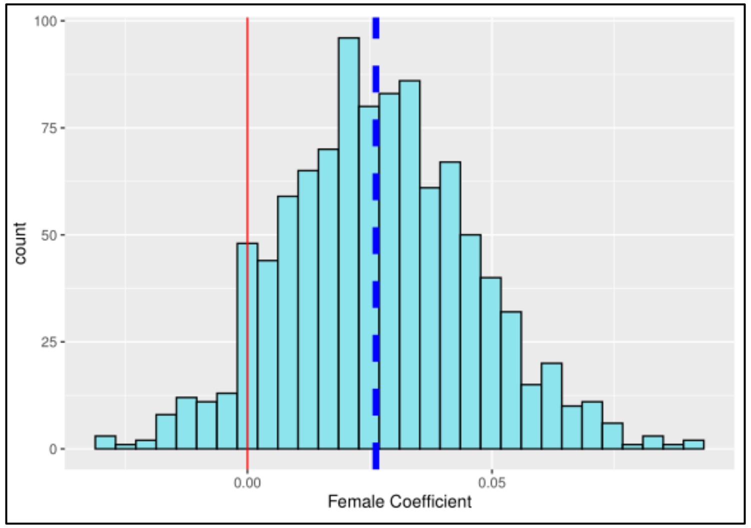
- The model could be run several times on bootstrapped samples
  - Bootstrapping involves sampling from replacement from the original dataset
  - The bootstrap samples have the same number of records
  - Each model run would result in different coefficients, since the dataset is different
- Evaluating the coefficients
  - The range of coefficients can be evaluated by variable
  - If the range of coefficients is narrow, it raises our confidence in statistical significance
  - If the range of coefficients is quite wide, it is a sign of model instability
  - Histograms can help visualize the range and distribution of coefficients
    - Narrower histograms with tall peaks are preferable
    - Variables where the histogram crosses over the 0 line should be further scrutinized



Bootstrapping

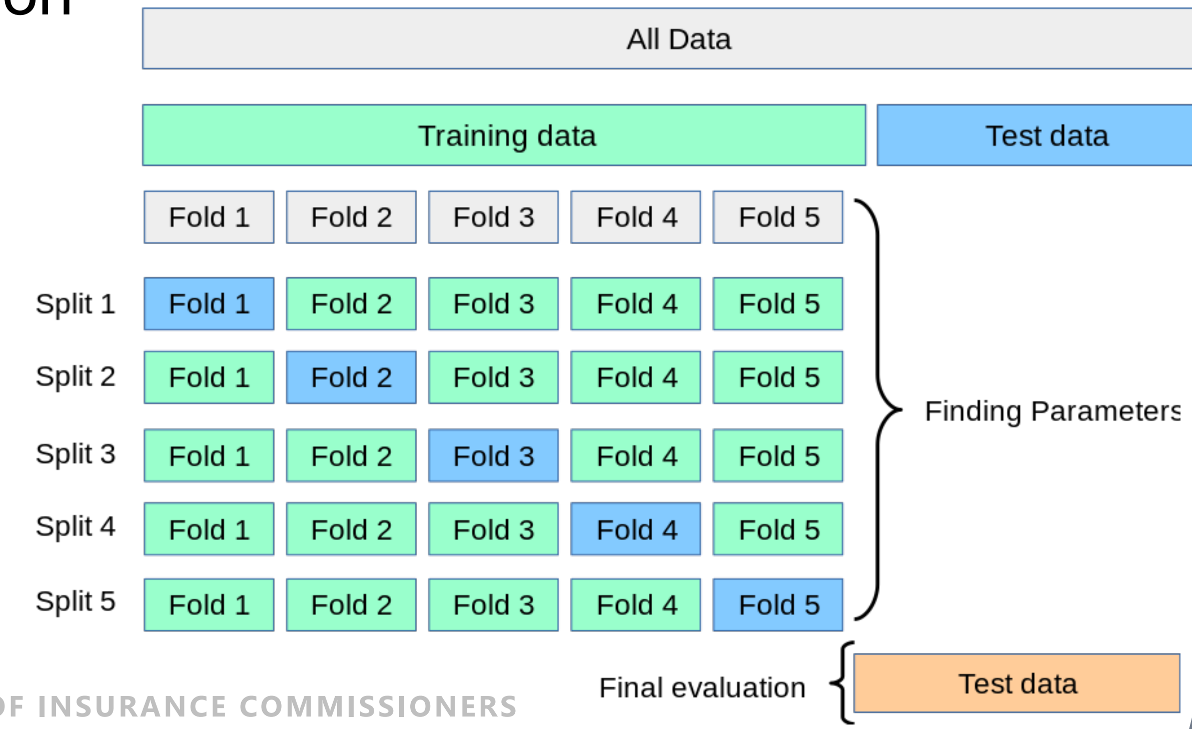
- Example
  - Elastic Net model was built
    - Glmnet package in R does not produce p-values
    - Instead, the same model was run 1,000 times on bootstrapped data samples
    - Histograms were analyzed to determine variability of coefficients by variable

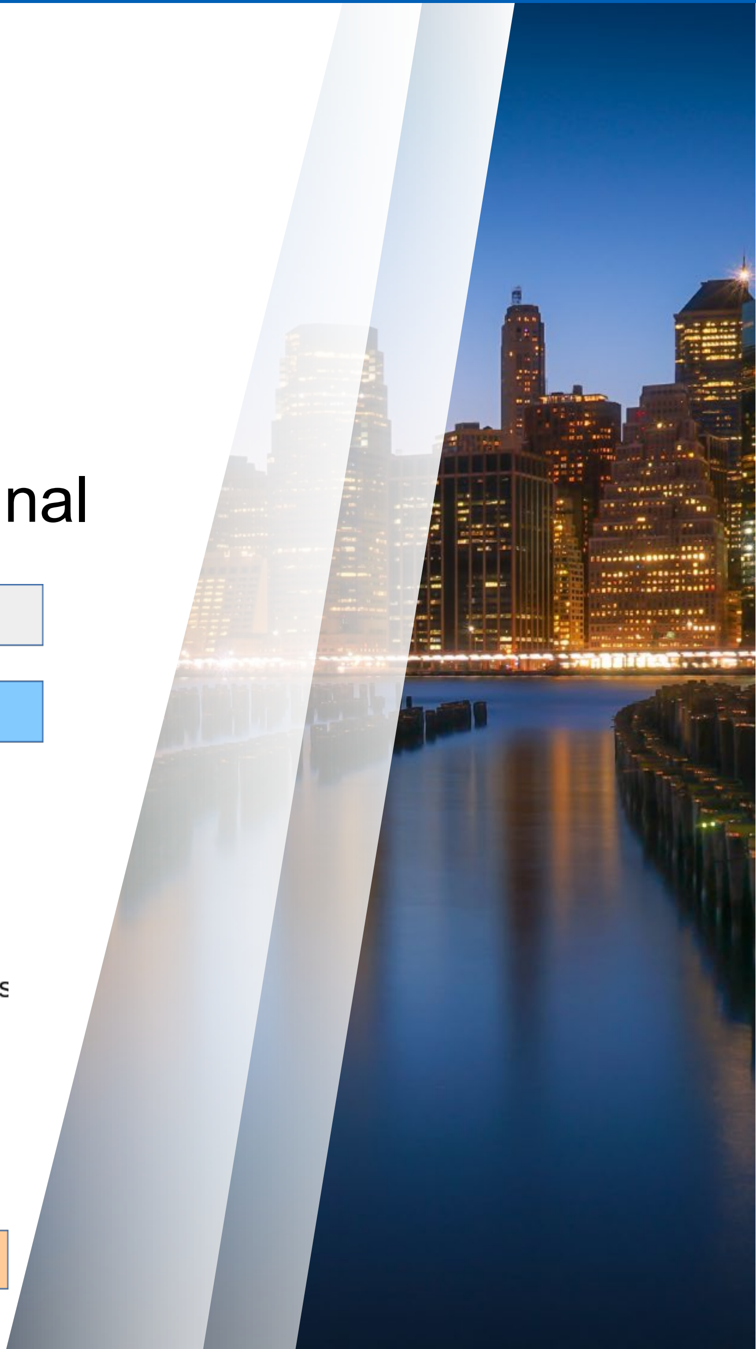




Cross Fold Validation

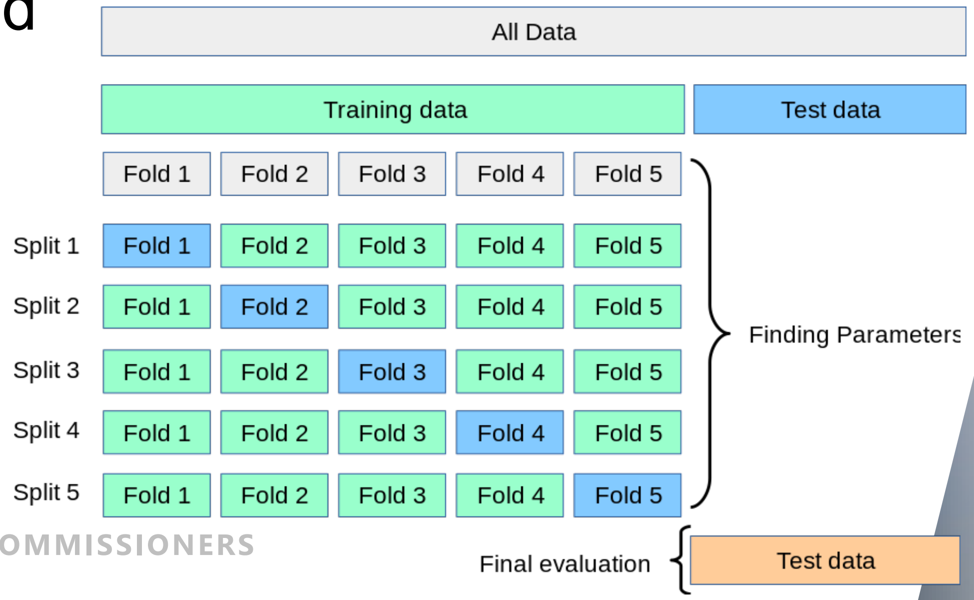
- K fold validation is a common cross fold validation type
- Training data is broken up into k folds
- Ideally, the modeler still has a true holdout dataset for final model validation





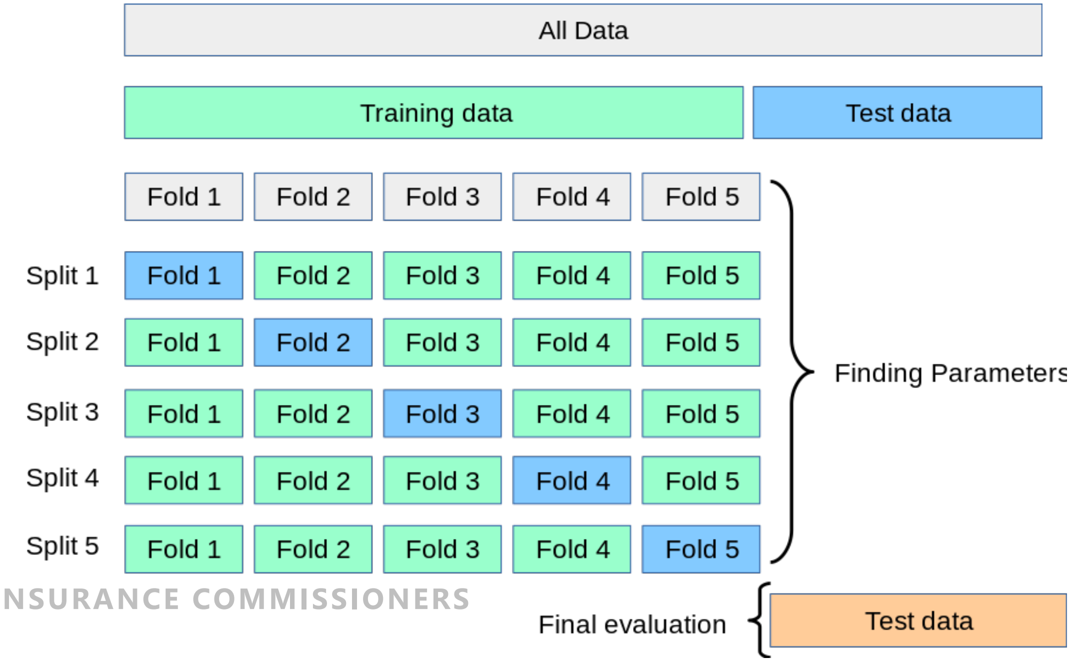
Cross Fold Validation

- The model is trained k times
- The predictions for a specific fold are based on a model trained with all other folds
- Each time the model is trained, a set of coefficients is determined
- The modeler may need to specify that they want each fold's coefficients to be saved



Cross Fold Validation

- Often the final model is run using 100% of the training data (including all folds)
- Companies often just provide coefficients associated with the final run
- However, reviewing the coefficients from the k folds may be useful



Cross Fold Validation

- The model reviewer can ask for the coefficients from each fold
  - If k fold validation was used, there are k different sets of coefficients
  - Unsure of the ideal k value
    - Small k values mean there are less sets of coefficients to analyze
    - Large k values mean that each model has a larger share of overlapping training data
  - Each model run would result in different coefficients, since the folds in training are different
- Evaluating the coefficients by fold
  - The range of coefficients can be evaluated by variable
  - If the range of coefficients is narrow, it raises our confidence in statistical significance
  - If the range of coefficients is quite wide, it is a sign of model instability
  - Histograms can help visualize the range and distribution of coefficients
    - Narrower histograms with tall peaks are preferable
    - Variables where the histogram crosses over the 0 line should be further scrutinized

Cross Fold Validation

- Example
  - Suppose we build an Elastic Net Model
    - Glmnet package in R does not produce p-values
    - Instead, the same model was run on 5 different folds
    - Consistency across folds can be analyzed

		1	2	3	4	5	Full Dateset
Usage	All Trips	0.530	0.518	0.560	0.755	0.690	0.618
	Professional	0.233	0.240	0.256	0.252	0.213	0.239
	Retired	-0.123	-0.134	-0.133	-0.152	-0.148	(0.138)
	Work Private			*Base*			
Gender	Female	0.039	(0.002)	0.034	0.020	0.041	0.026
	Male			*Base*			
Driver Age	16 - 20	0.132	0.124	0.185	0.080	0.271	0.161
	21 - 30	(0.037)	0.042	0.009	0.001	0.042	0.011
	31 - 40			Base			
	41 - 50	(0.039)	(0.043)	(0.042)	(0.022)	(0.013)	(0.032)
	51 - 60	0.028	0.009	0.005	0.049	0.043	0.027
	61+	0.067	0.063	0.063	0.104	0.112	0.082
Vehicle Age	0 - 5	0.116	0.132	0.109	0.109	0.114	0.116
	6 - 10			*Base*			
	11+	(0.400)	(0.385)	(0.379)	(0.404)	(0.360)	(0.386)
Vehicle Din	0 - 50	(0.606)	(0.607)	(0.534)	(0.631)	(0.635)	(0.601)
	51 - 100			*Base*			
	101 - 150	0.211	0.213	0.198	0.220	0.220	0.212
	151+	0.271	0.227	0.250	0.275	0.241	0.253

GLM Reference Model

- GLMs provide p-values in most software
- A GLM could be built which is as similar as possible to the model in question
  - This is probably more appropriate when the model in question is still some type of linear model (Lasso, ridge, elastic net)
- Consider the GLM provided p-values a reasonable approximation for the model in question
  - P-values from the GLM may be a little underestimated
- The modeler should describe why their model type is preferable to a GLM for their modeling purpose.
  - Once they have a similar GLM, they should describe why they favor the other model
  - Why not use Lasso or Elastic Net for variable selection, but run a GLM on the final features?
- If the coefficients are radically different in the reference GLM, the GLM p-values may not be as relevant

GLM Reference Model

- Example
  - Elastic Net model was built
  - GLM model was built with the same variables
  - The coefficients are compared side by side
  - Low p-values from the GLM suggest the variables should be significant

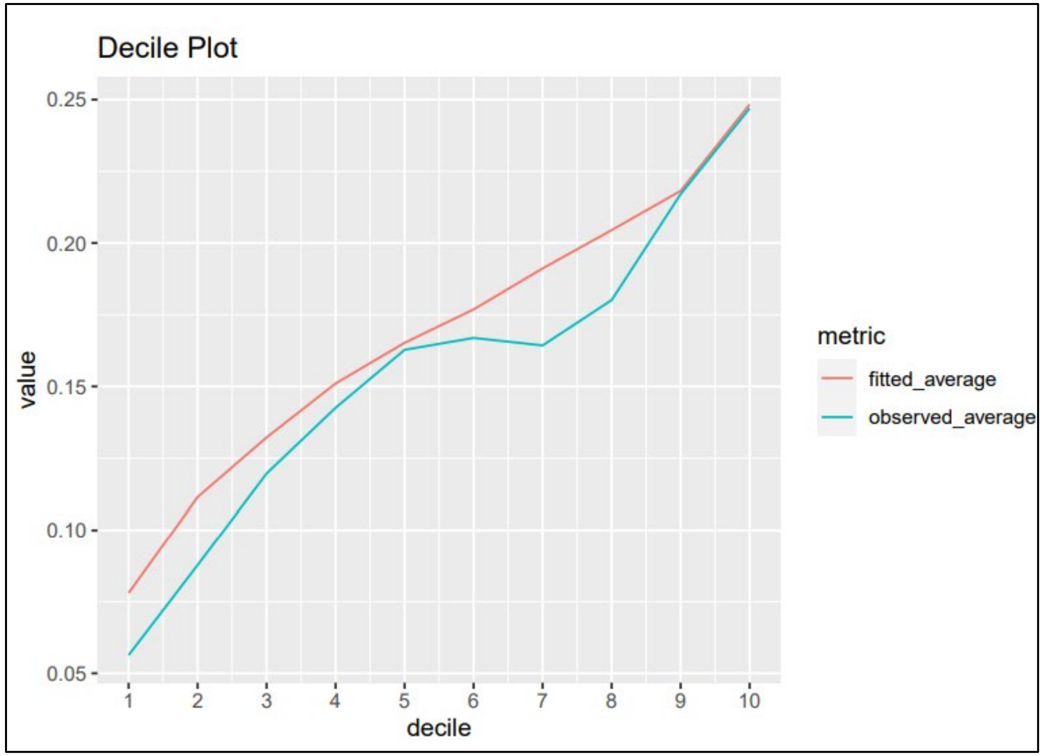
		Elastic Net	Reference GLM	GLM p-value
Usage	All Trips	0.618	0.622	< 0.001
	Professional	0.239	0.239	0.002
	Retired	(0.138)	(0.142)	< 0.001
	Work Private	*Base*	*Base*	
Gender	Female	0.026	0.027	0.157
	Male	*Base*	*Base*	
Driver Age	16 - 20	0.161	0.170	0.398
	21 - 30	0.011	0.014	0.769
	31 - 40	*Base*	*Base*	
	41 - 50	(0.032)	(0.031)	0.320
	51 - 60	0.027	0.029	0.329
	61+	0.082	0.087	0.016
Vehicle Age	0 - 5	0.116	0.116	< 0.001
	6 - 10	*Base*	*Base*	
	11+	(0.386)	(0.386)	< 0.001
Vehicle Din	0 - 50	(0.601)	(0.606)	< 0.001
	51 - 100	*Base*	*Base*	
	101 - 150	0.212	0.213	< 0.001
	151+	0.253	0.255	< 0.001

Comparison of Alternatives

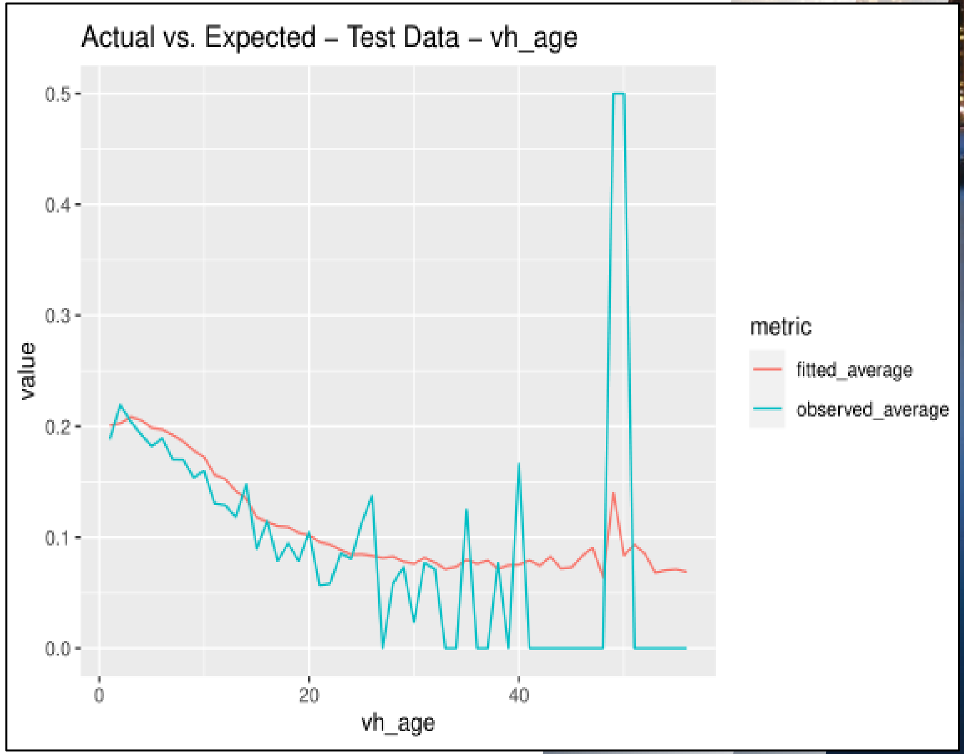
- Bootstrapping
  - Can provide a large distribution of coefficients
  - May be impractical for large datasets due to model run time
- K Fold Validation
  - Typically provides a much smaller distribution of coefficients
  - Often requires the modeler to change programming to save coefficients from each fold
  - Takes less time than the bootstrapping approach since there are less model runs
- GLM Reference Model
  - Less appropriate for non-linear models
  - The p-values may not be relevant if the beta coefficients are radically different from the model in question

Applicable to ALL Supervised Learning

- Quantile Plots on Holdout Data
  - Compare fitted average to observed average by quantile

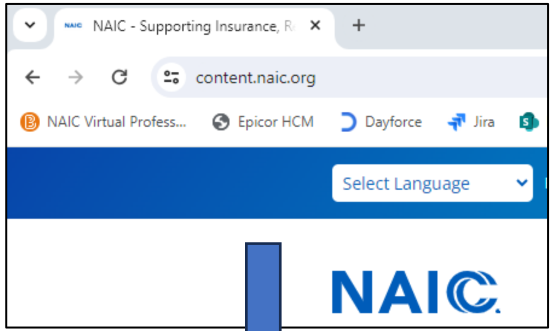


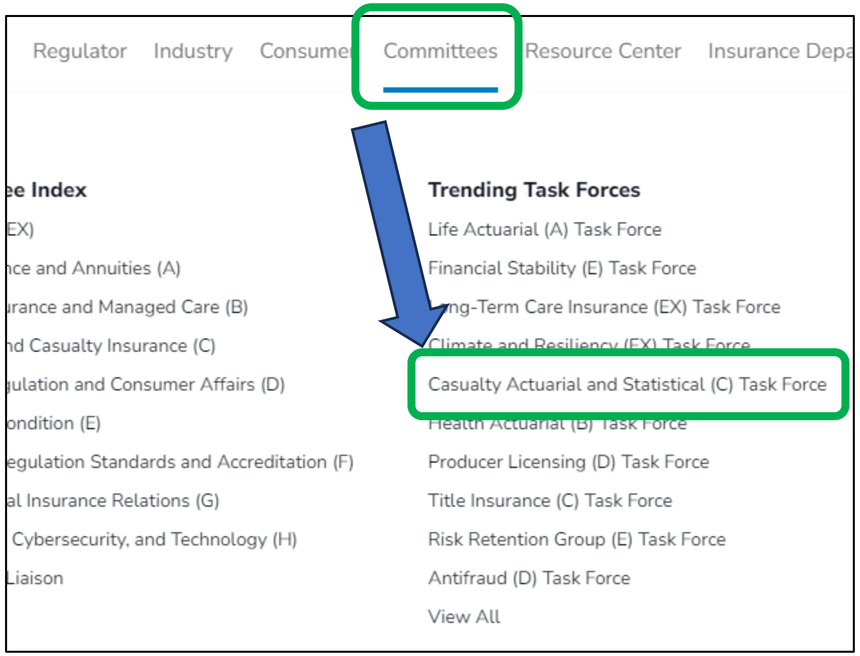
- Actual vs. Expected plots on Holdout Data
  - Separate plots by variable
  - Demonstrate fit across levels

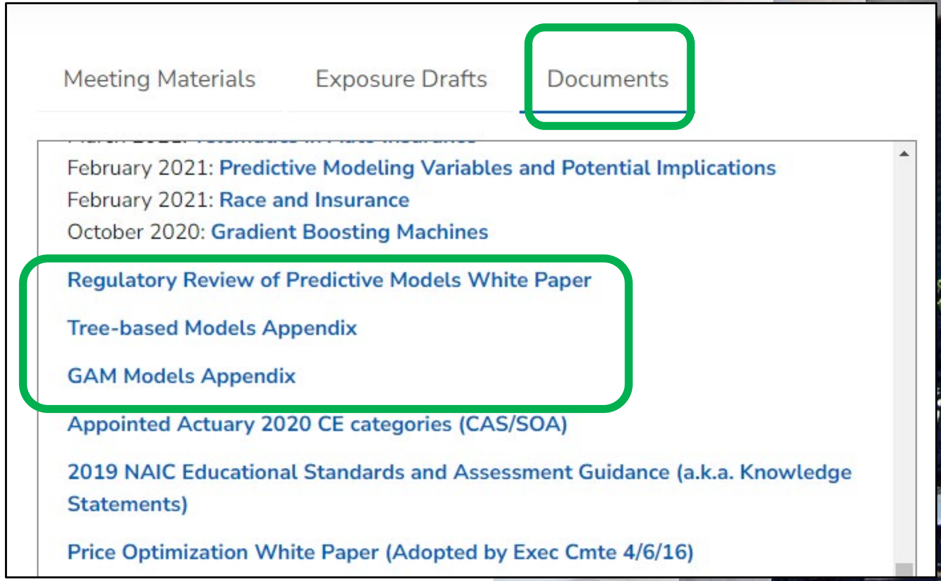


CASTF White Paper and Appendices

www.naic.org

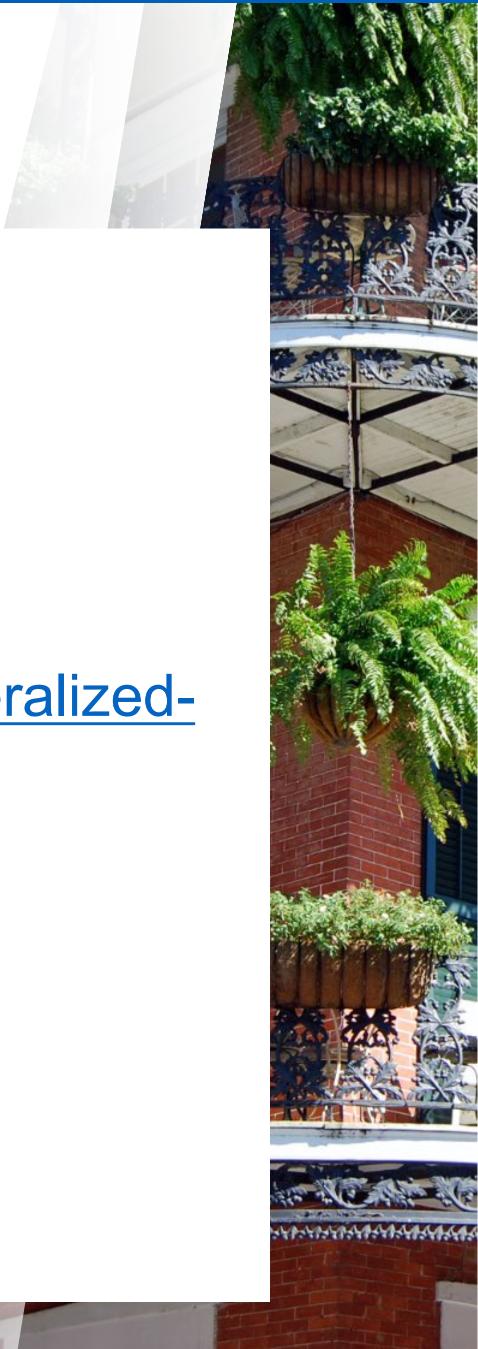






GAM References

- June 2021 Book Club: Generalized Additive Models GAM
  - https://www.youtube.com/watch?v=F1fMKy4fMIk
- April 2021 Book Club: From GLMs to GAMs
  - https://www.youtube.com/watch?v=vRbHqbNINx8
- DataCamp R coding course: Nonlinear Modeling with GAMs in R
  - https://app.datacamp.com/learn/courses/nonlinear-modeling-with-generalized-additive-models-gams-in-r



Tree Based Model References

- Basic Decision Tree Terminology
  - https://medium.datadriveninvestor.com/the-basics-of-decision-trees-e5837cc2aba7
- Theoretical Introduction to Random Forest
  - Introduction to Statistical Learning (Chapter 8 – 8.2.2)
  - https://web.stanford.edu/~hastie/ISLRv2_website.pdf
- Interpretable Machine Learning (Variable Importance and Interpretability Plots)
  - https://us.milliman.com/-/media/milliman/pdfs/2021-articles/4-2-21-interpretable-machine-learning.ashx
  - Book Club Presentation: https://www.youtube.com/watch?v=-yMdTAIkewk
- Tree-Based Models Book Club
  - https://youtu.be/6UCbpAt4r9M

Other Penalized Regression References

- L1 and L2 Penalized Regression Models
  - https://cran.r-project.org/web/packages/penalized/vignettes/penalized.pdf
- October 2022 Book Club: P-values and Alternatives
  - https://www.youtube.com/watch?v=_V_z6f4L1qw
- October 2023 Book Club: Derivative Lasso and Lasso Credibility
  - https://www.youtube.com/watch?v=_V_z6f4L1qw
- Akur8 White Papers
  - https://www.akur8.com/resources/white-papers
- AGLM: A Hybrid Modeling Method of GLM and Data Science Techniques
  - https://www.institutdesactuaires.com/global/gene/link.php?doc_id=16273&fg=1
- Cross-validation: Evaluating Estimator Performance
  - https://scikit-learn.org/stable/modules/cross_validation.html