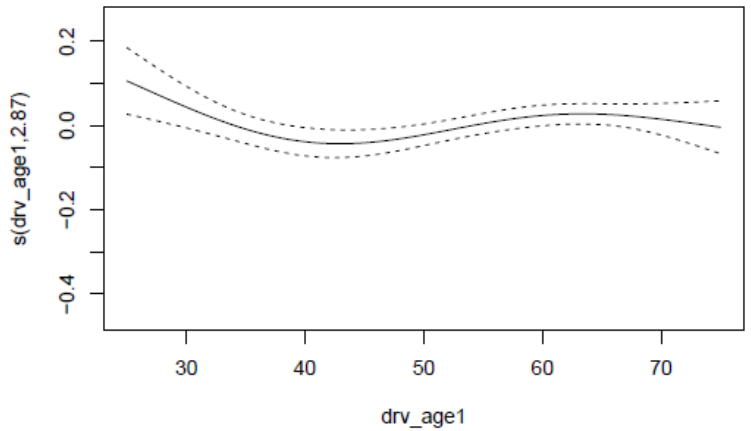
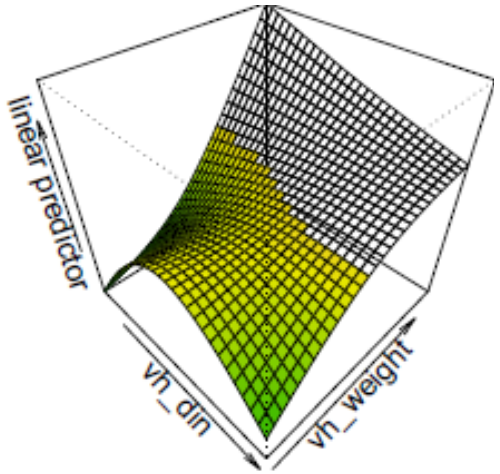
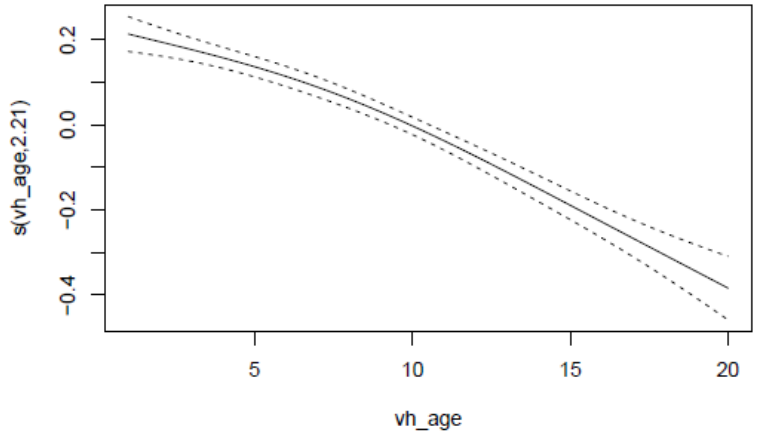






Generalized Additive Models

Sam Kloese, ACAS, CSPA
P/C Rate Modeling Actuary
December 2, 2022

NAIC NATIONAL ASSOCIATION OF INSURANCE COMMISSIONERS

Image from Introduction to Boosted Trees – The Official Blog of BigML.com

Introduction

- GLMs are industry standard
- The CASTF White Paper for Predictive Models is focused primarily on GLM's
- New Appendix for Tree Based Models has already been adopted
- New proposed appendix will cover GAMs
  - GAMs are similar to GLMs with just a few differences
  - The original GLM appendix was used as a starting point

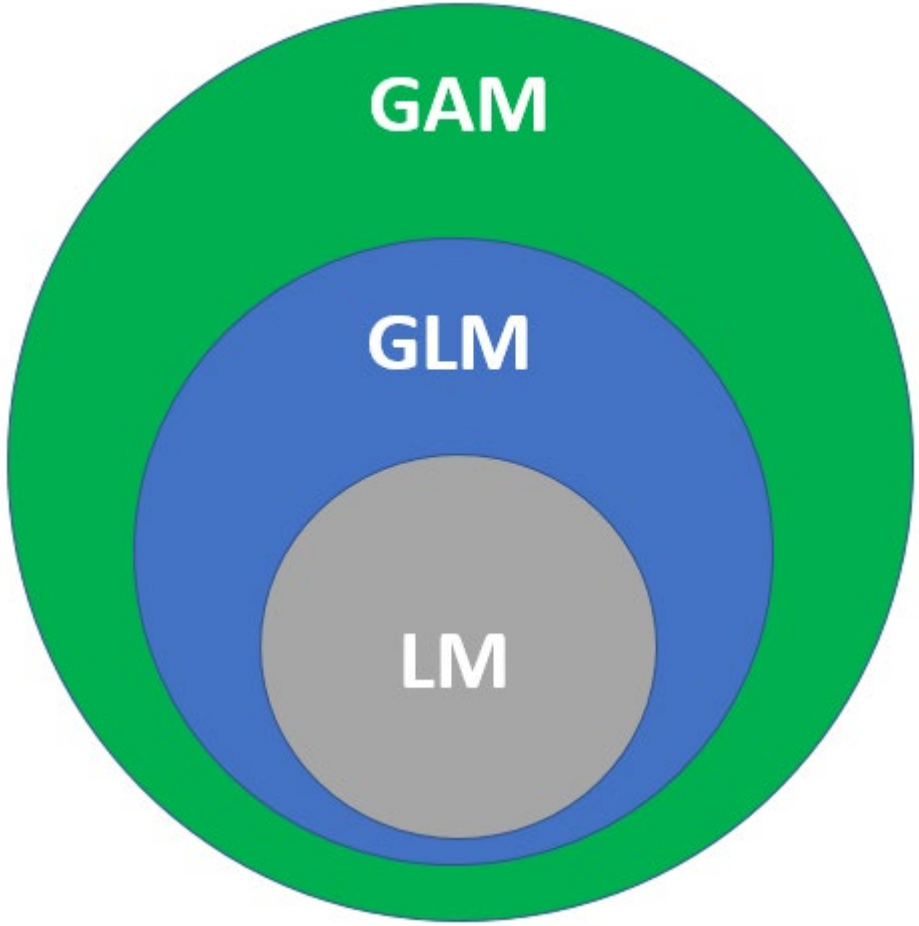
Similarity to GLM's

- 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…)
  - Link Function defines the relationship between the linear predictor and the mean (log link, logistic link…)
  - Offset terms can be added
  - Records can be weighted (exposures in a frequency model…)

```
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 GLM's

- 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
  - A LM is a special case of GLM
    - Distribution: Normal
    - Link Function: Identity
  - A GLM is a special case of GAM
    - No smoothed terms

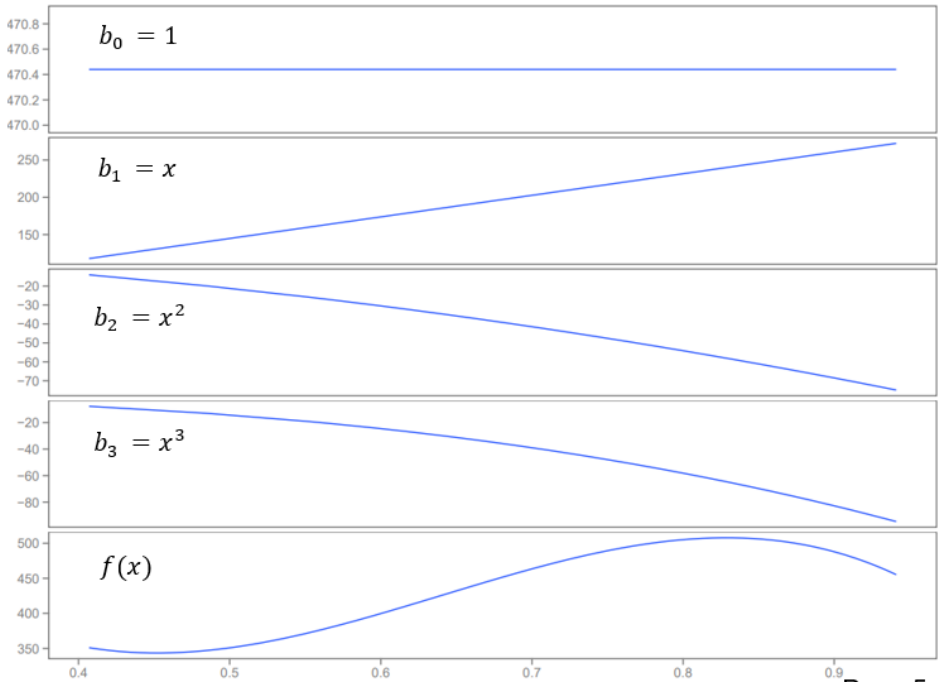


Smooth functions

- 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

Polynomial Basis Example

$$f(x) = \beta_0 1 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$



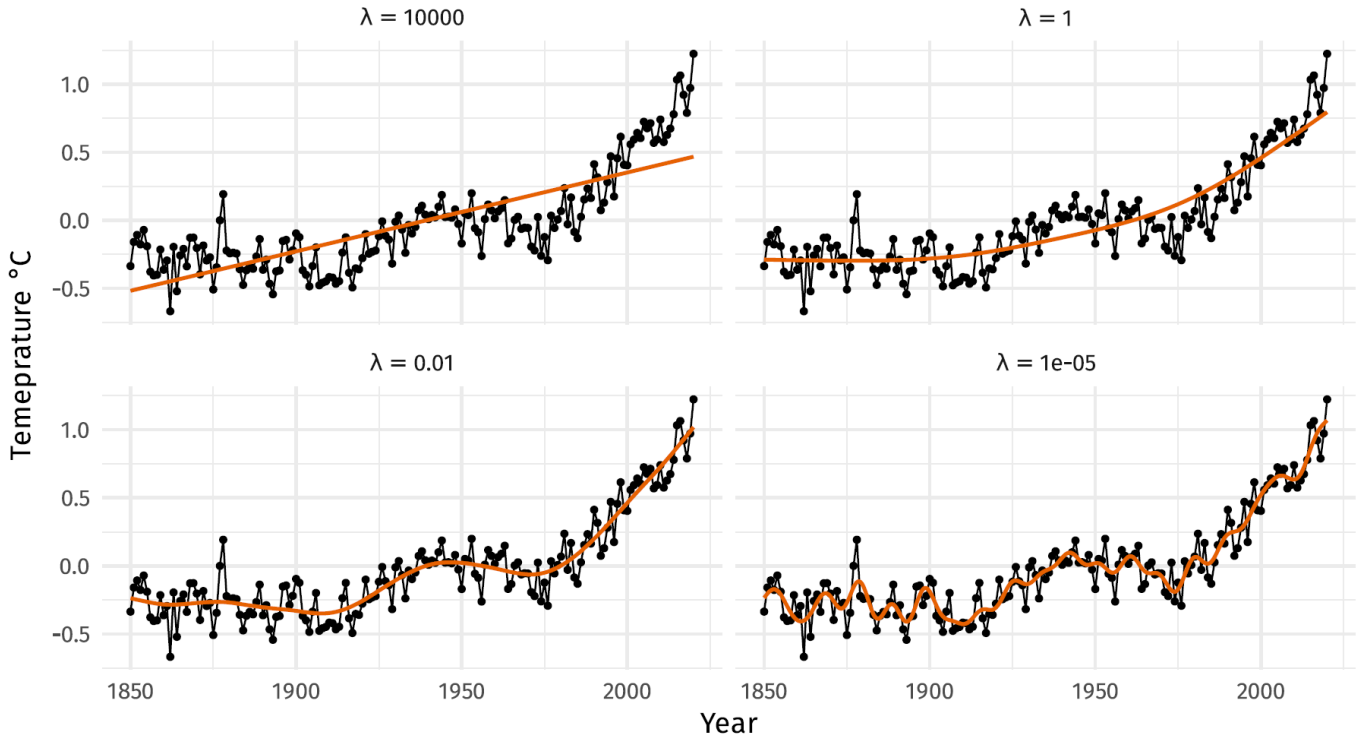
GAM is a type of Penalized Regression

- Other forms of penalized regression
  - Lasso, Ridge, Elastic Net
- 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

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

GAM is a type of Penalized Regression

HadCRUT4 time series



HadCRUT4 is a global temperature dataset, providing gridded temperature anomalies across the world as well as averages for the hemispheres and the globe as a whole.

The smaller the λ the wigglier the fit.

The modeler sets smooths and related k values, the software typically chooses λ.

Complications from Smoothed Terms

- Smoothed terms have multiple beta coefficients
  - Relationship no longer summarized within 1 single number
  - The impact of the smoothed term is hard to interpret without plots
- P-values are less straightforward
  - The calculation changes for penalized regression methods
  - The mgcv package provides p-values, but they are approximate
- Smoothed terms introduce the risk of “concurvity”
  - Concurvity is similar to the concept of collinearity in the parametric (non-smoothed) terms
  - Concurvity is when the smoothed terms move together

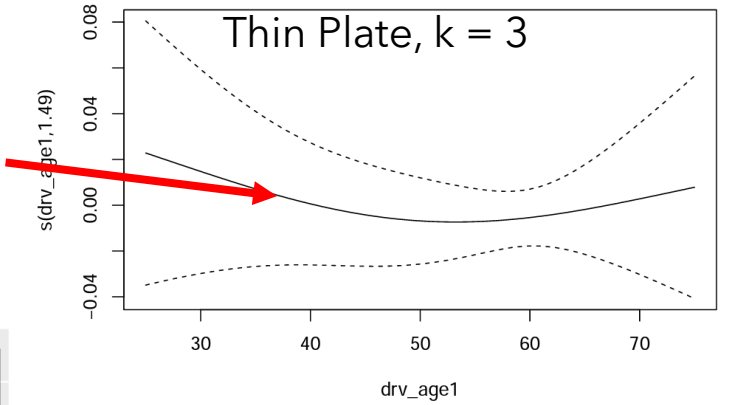
Recommendations for Smoothed Terms

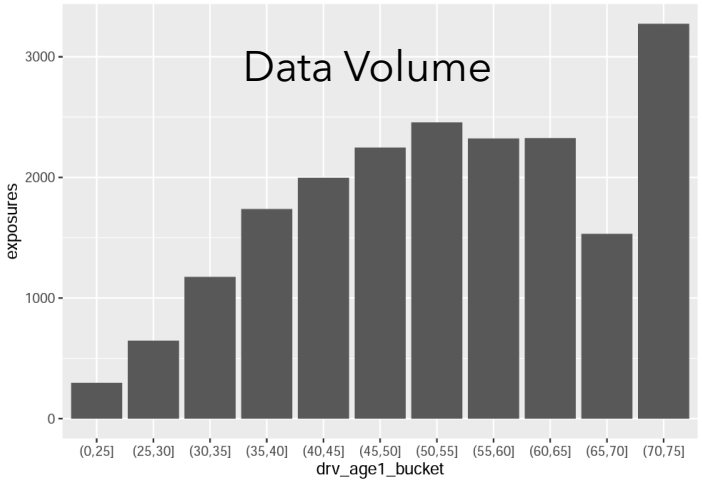
1. Review plots for each smoothed term
2. Review approximate p-value for each smoothed term
3. Review concurvity metrics

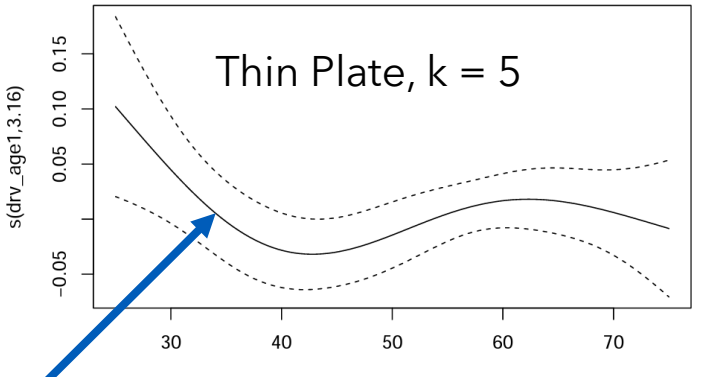
Recommendations for Reviewing Plots

- 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

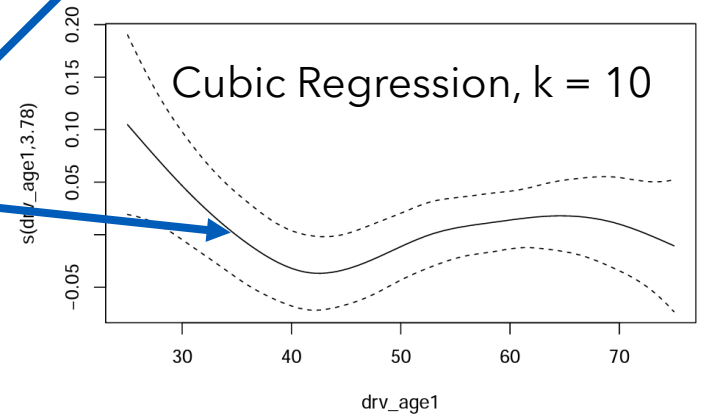
Extremely wide confidence intervals
Fails horizontal line test







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



Recommendations for Reviewing 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
```

Recommendations for Reviewing Concurvity

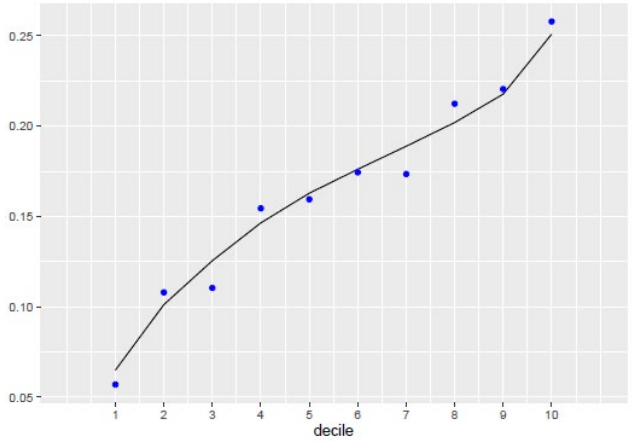
- Mgcv provides 3 versions of concurvity metrics: worst, observed, estimate
- Worst is the most pessimistic view
- Rule of thumb, a worst concurvity > 0.8 is too high for a smoothed term

```
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
```

GLM Tests that are still applicable

- Overall model fit tests
  - Quantile Plots (Lift Curves)
  - Lorenz Curves (Measures segmentation power)
- Significance by term
  - (Approximate) P-value
  - Confidence Intervals (within plots)
  - AIC* would ideally decrease after each term added
  - F nested model test
  - Residual Plots
  - Univariate Observed Average vs. Predicted Average



Formula	Predictors	AIC
~ pol_coverage	1	75,573
~ pol_coverage + pol_usage	2	75,422
~ pol_coverage + pol_usage + s(drv_age1, k = 4)	3	75,406
~ pol_coverage + pol_usage + s(drv_age1, k = 4) + s(vh_age, k = 4)	4	75,118
~ pol_coverage + pol_usage + s(drv_age1, k = 4) + s(vh_age, k = 4) + te(vh_din, vh_weight, k = 3)	5	74,955

**The formula for AIC is different from a GLM, but is still readily available in R*

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