PINNACLE
ACTUARIAL RESOURCES

From GLMs to GAMs

April 27, 2021

Radost Roumenova Wenman, FCAS, MAAA, CSPA
Consulting Actuary

Introduction

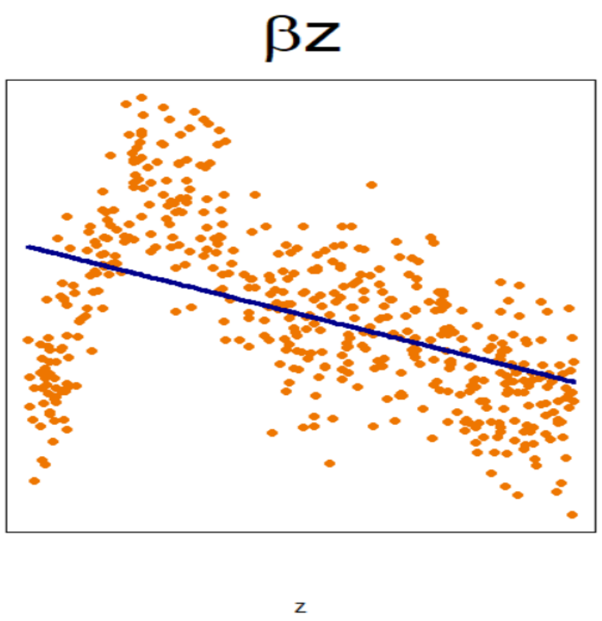
Generalized Linear Models **GLM**

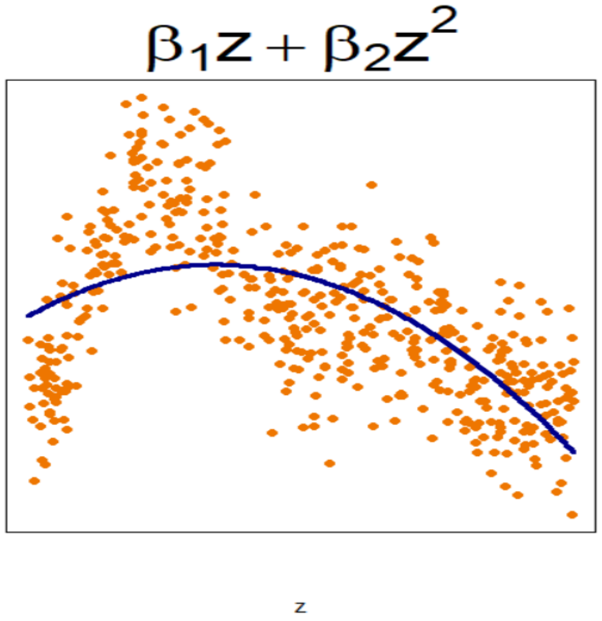
$$\hat{g}\left(E(y)\right) = \boldsymbol{X}^T\boldsymbol{\beta}$$

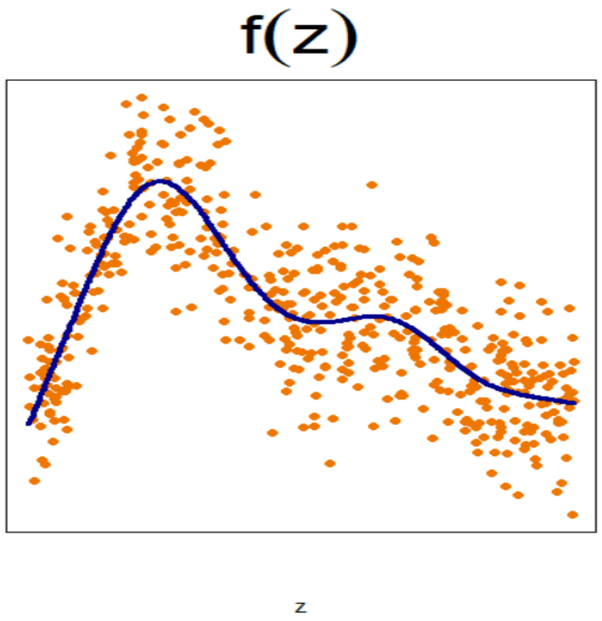
Generalized Additive Models **GAM**

$$\hat{g}\left(E(y)\right) = \boldsymbol{X}^T\boldsymbol{\beta} + f(z)$$

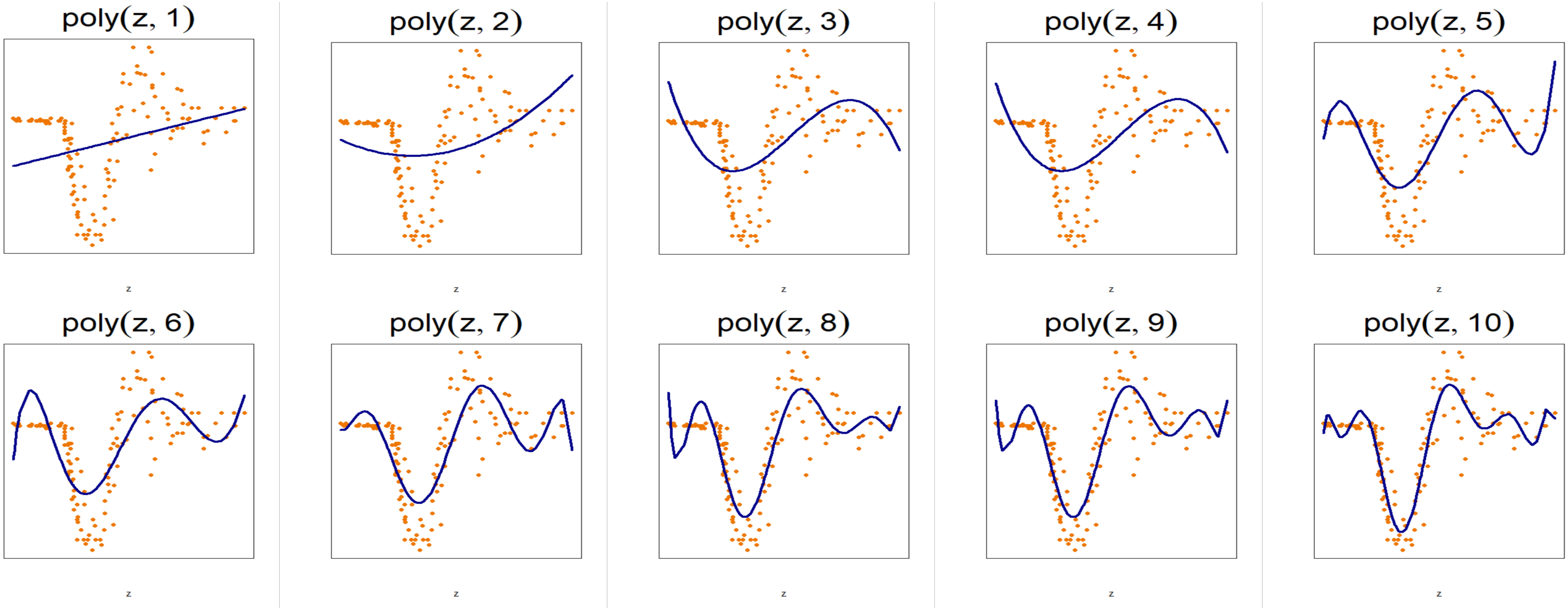
$$= \textbf{GLM} + f(z)$$







How about Polynomial Fits?



Background

Trevor Hastie and Robert Tibshirani (1986)

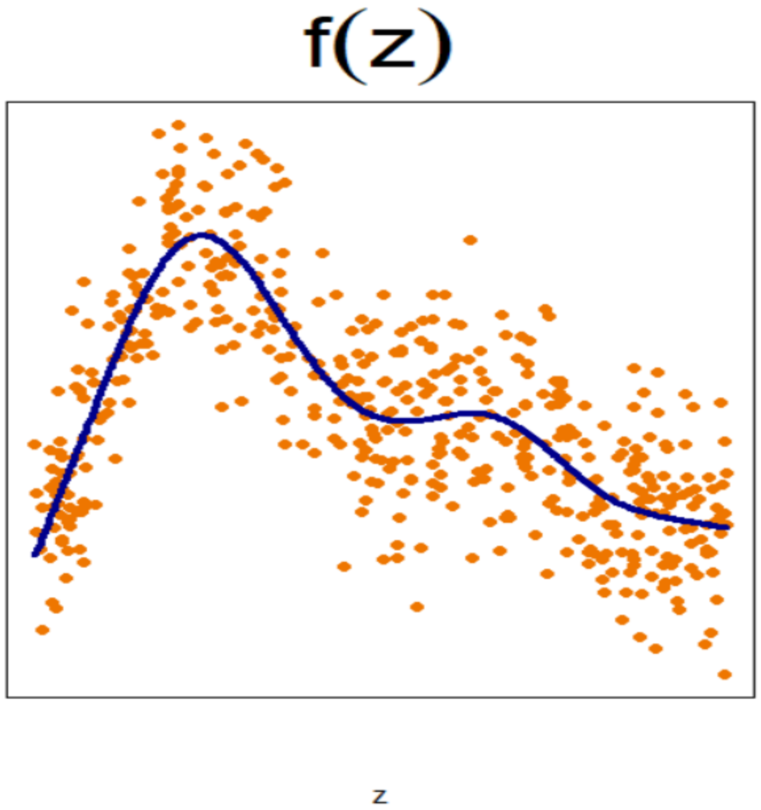
Replace the linear predictor with an "additive" predictor

Utilizes smooth functions

Useful in uncovering nonlinear effects

Completely automatic

"No detective work is needed…"



GAMs vs. GLMs

General Linear Model (GLM)

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p + \epsilon$$

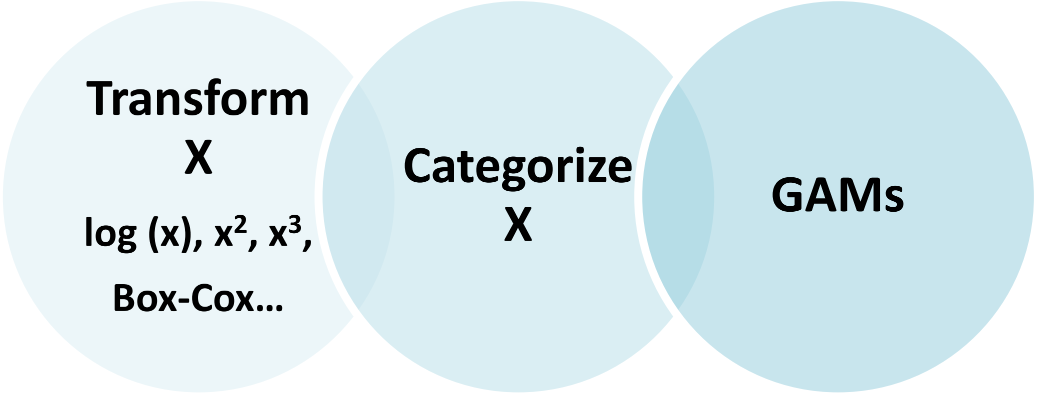
Generalized Linear Models (GLMs)

$$g(E_Y(y|x)) = \beta_0 + \beta_1 x_1 + \ldots \beta_p x_p$$

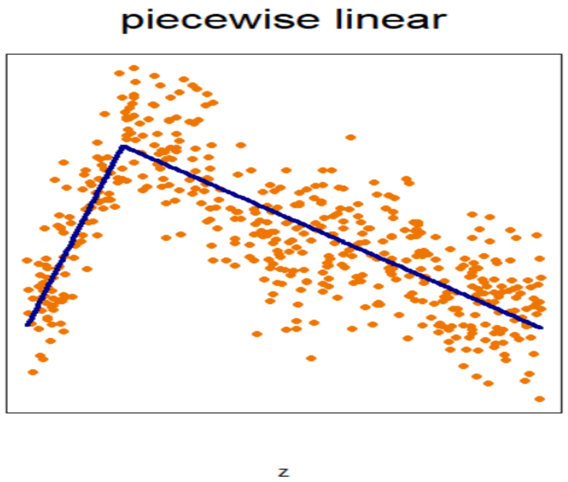
Generalized Additive Models (GAMs)

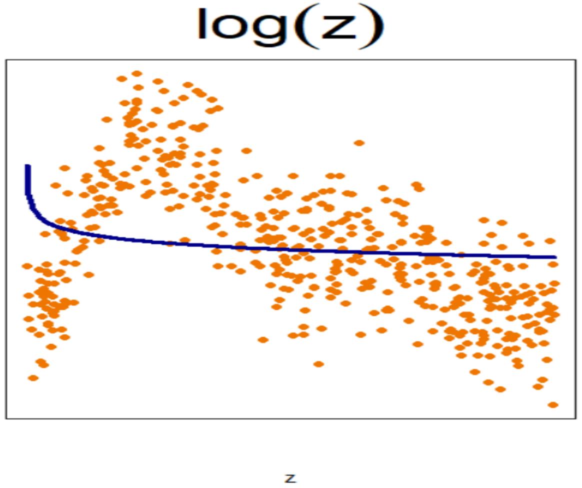
$$g(E_Y(y|x)) = \beta_0 + f_1(x_1) + f_2(x_2) + \ldots + f_p(x_p)$$

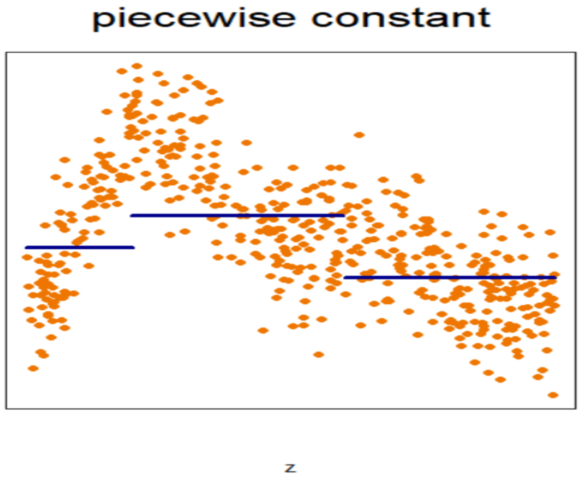
How Can We Address Nonlinearity?

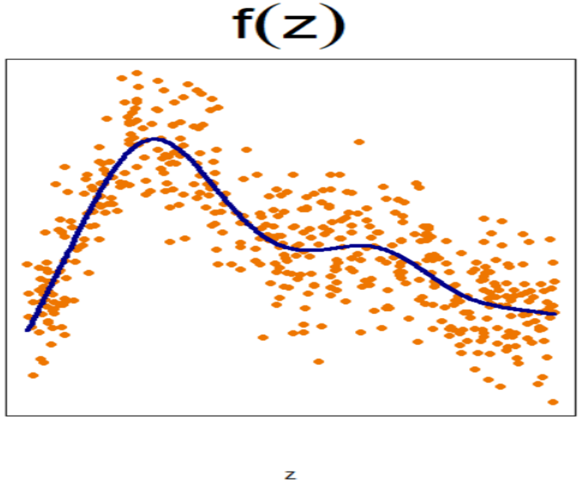


Modeling Nonlinearity

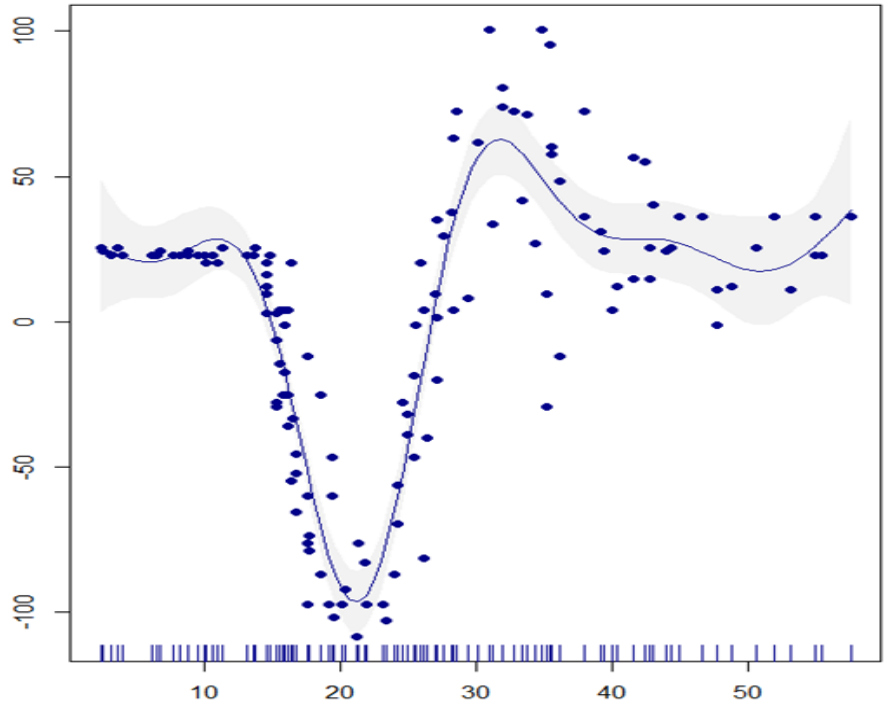


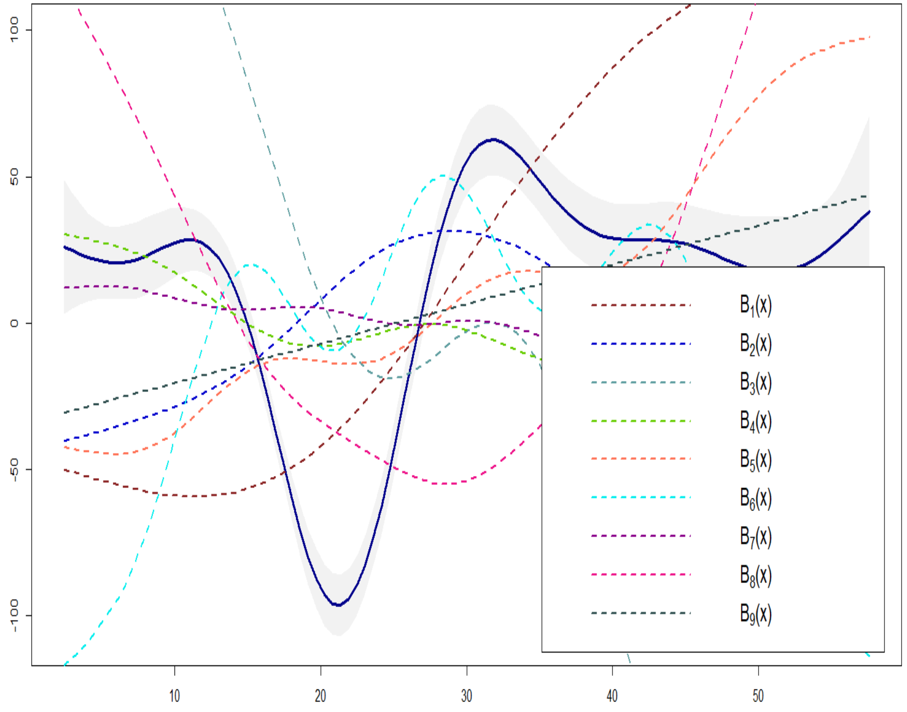






Basis Functions

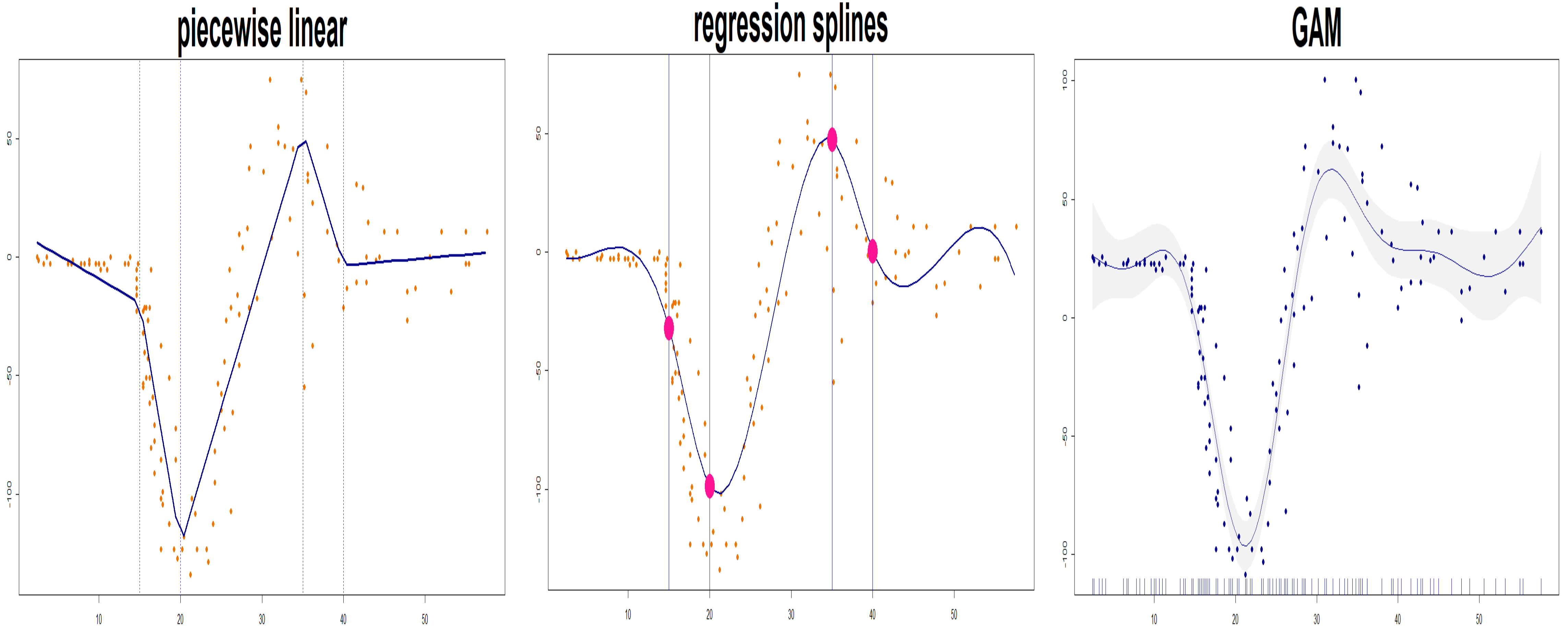




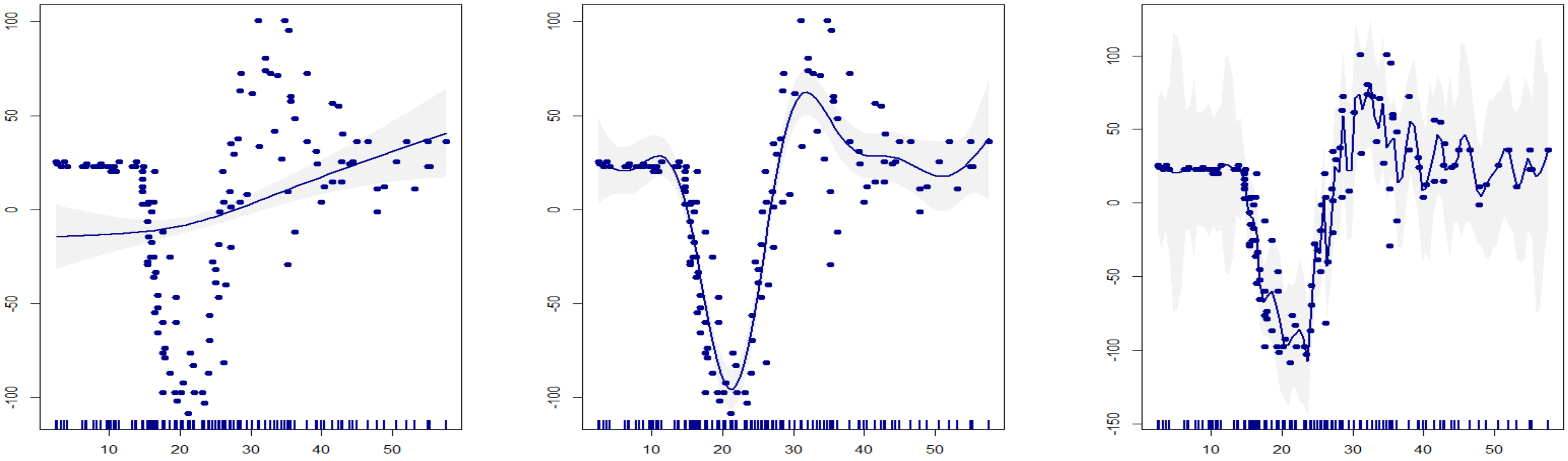
$$g[E(Y|X)] = b_0 + b_1 X^1 + b_2 X^2 + \cdots + b_9 X^9$$

$$\mathbf{f}(\boldsymbol{X}) = g[E(Y|X)] = b_0 + b_1 \mathrm{B}_1(\mathrm{X}) + b_2 \mathrm{B}_2(\mathrm{X}) + \cdots + b_9 \mathrm{B}_9(\mathrm{X})$$

Regression Splines



Smoothing Splines and the Bias-Variance Trade-off



$$\sum_{i=1}^{n} \left(y_i - b(x_i)\right)^2 + \lambda \int \left(b''(x)\right)^2 \mathrm{d}x$$

GAM – Example with Poisson

$\log(\mu_i) = \log(n_i) + \eta_i = \log(n_i) + \beta z_i$ ➡ **GLM**

$\log(\mu_i) = \log(n_i) + \eta_i = \log(n_i) + f(z_i)$ ➡ **GAM**

$$l(z,\mu) = \sum_{i=1}^{n} [z_i \log(\mu_i) - \mu_i - \log(z_i!)]$$

$$= \sum_{i=1}^{n} [z_i \, \mathrm{f}(z_i) - \exp(\mathrm{f}(z_i)) - \log(z_i!)]$$

$$= \sum_{i=1}^{n} [z_i \, \mathrm{f}(z_i) - \exp(\mathrm{f}(z_i)) - \log(z_i!)] - \frac{1}{2}\lambda \int [f''(z)]^2 \, \mathrm{dz}$$

Flexibility of GAMs

$$g(E_Y(y|x)) = \beta_0 + f_1(x_1) + f_2(x_2) + \ldots + f_p(x_p)$$

- Multiple predictors
- Mixture of smoothing splines, linear terms, and nominal variables
- Smooth interactions

GAM Summary Output

Family: gaussian   Link function: identity

parametric coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -25.546   1.951   -13.1   <2e-16 ***

coef(gam_mod) – smooth terms:

s(z).1	s(z).2	s(z).3	s(z).4	s(z).5
-63.718	43.476	-110.350	-22.181	35.034

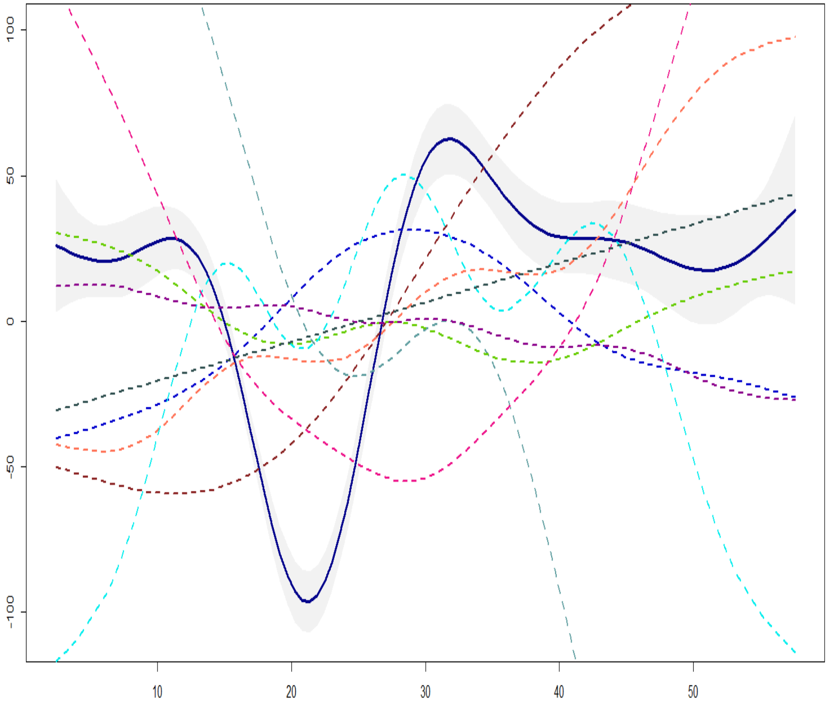
s(z).6	s(z).7	s(z).8	s(z).9
93.176	9.283	-111.661	17.603

Approximate significance of smooth terms:

	edf	F	p-value
s(z)	8.693	53.52	**<2e-16 *****

R-sq.(adj) = 0.783   Deviance explained = 79.8%

GCV = 545.78   Scale est. = 506   n = 133

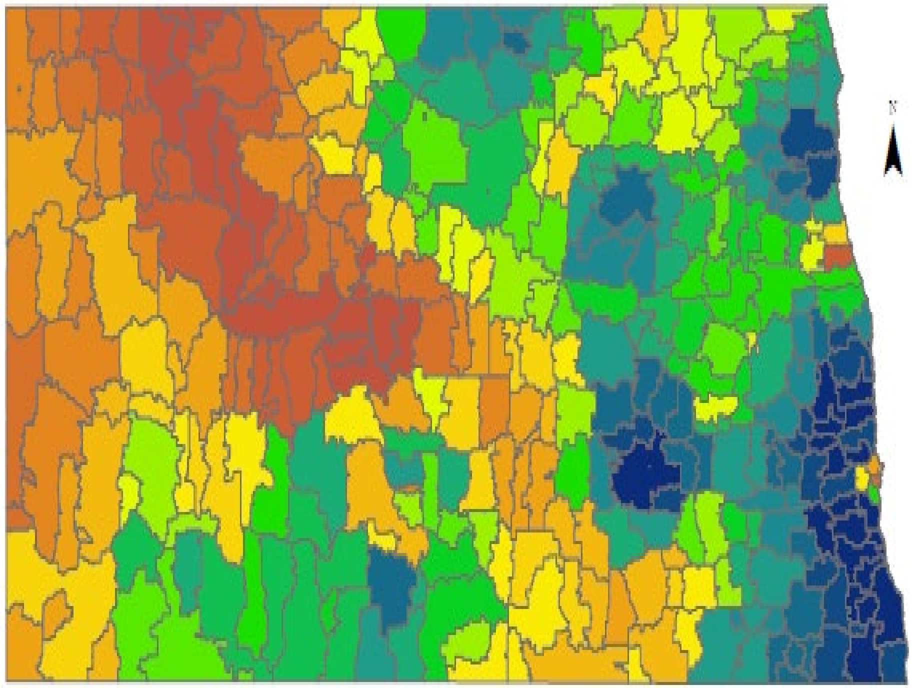


Hypothesis Testing, GCV, AIC, Stepwise Variable Selection, Shrinkage

Insurance Application of GAMs – Geospatial Smoothing

Including geographic territories directly in a GLM is generally not feasible!

- Popular technique – smoothing and clustering
  - zero exposure?
  - homogeneous?
  - clustering method?
- Alternative technique – GAM
  - directly applies spatial smoothing
  - can use longitude and latitude



GAM Approach to Modeling Geolocation Data

- Method 1 (two-step)
  - include non-geographic variables as predictors in a GLM
  - extract the GLM residuals
  - use GAM to regress the GLM residuals on **f(longitude, latitude)**
- Method 2 (two-step)
  - include non-geographic variables as predictors in a GLM
  - extract the GLM linear predictor
  - Use the GLM linear predictor as an offset in a GAM only with **f(longitude, latitude)**
- Method 3 (one-step)
  - include all variables, including geolocation, as predictors in a GAM

In a Nutshell…

GAMs = Penalized GLMs!

Recommended References

- Hastie, T., Tibshirani, R. (1990). *Generalized Additive Models*, Chapman & Hall/CRC.
- Wood, S. (2017). *Generalized Additive Models: An Introduction with R*, Chapman & Hall/CRC.
- Fahrmeir, L., Kneib, T., Lang, L., Marx, B. (2013). *Regression: Models, Methods and Applications*, Springer.
- Klein, N., Denuit, M., Lang, S., and Kneib, T. (2014). *Nonlife Ratemaking and Risk Management with Bayesian Generalized Additive Models for Location, Scale, and Shape*. Insurance: Mathematics and Economics 55:225–49.
- http://www.variancejournal.org/issues/13-01/141.pdf
- https://www.soa.org/globalassets/assets/files/e-business/pd/events/2020/predictive-analytics-4-0/pd-2020-09-pas-session-006.pdf

Thank You

Radost R. Wenman, FCAS, MAAA, CSPA

415.692.0260

RWenman@PinnacleActuaries.com

PINNACLE
ACTUARIAL RESOURCES