A Brief Introduction to Generalized Linear Mixed Models and Generalized Additive Models

ERIC MANN, FCAS

URI KORN, FCAS

Roadmap

- Review of Linear Models and Generalized Linear Models
- Generalized Additive Models
  - Example
- Mixed Effect Models
  - Example
- Questions

Review of Linear Models

Classical Linear Model

- Response: $Y \sim N(X\beta, \sigma^2)$
- $X\beta$ is a linear function that describes how the expected values vary based on characteristics in the data
- Linear: $\beta_0 + \beta_1 X_1^2 + sin(\beta_2 X_2)$
- Non-linear: $\beta_1 X_1 e^{\beta_2 X_2}$
- Constant Variance

Generalized Linear Model

- Response: Poisson, Gamma, Binomial, etc.
- $Y \sim F(\pi, R)$
- Expected Value:
  $G(E[Y])^{-1} = X\beta$
- Variance is a function of expected value
- Responses are independent

Generalized Additive Models

- Linear predictor has a more general form

- $E\left(Y|X_1, X_2, \cdots X_p\right) = \alpha + f_1(X_1) + f_2(X_2) + \cdots + f_p\left(X_p\right)$

- $f_i(X_i)$ are non-parametric smoother functions

  - Smoothing Splines

  - Kernel Smoothers

  - Local Linear Regression

  - But can also be parametric functions, too

What Does That Mean in Real Life?
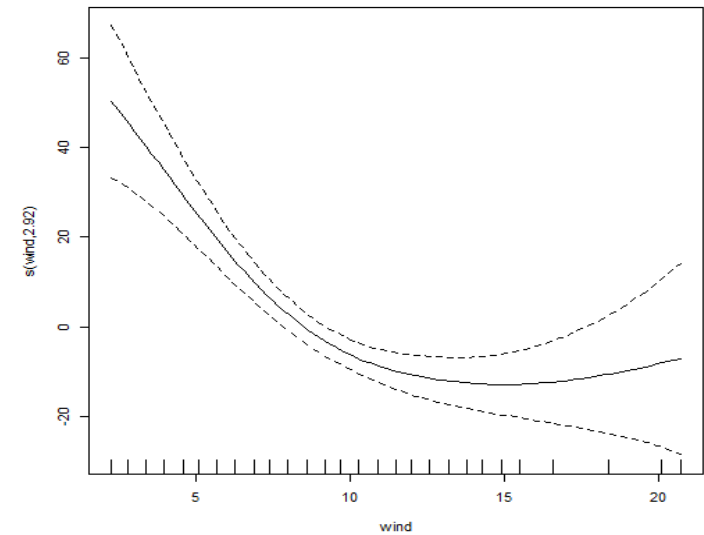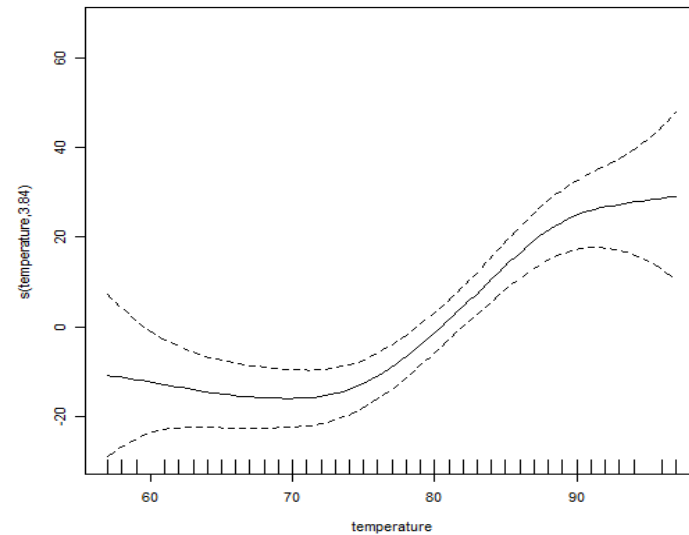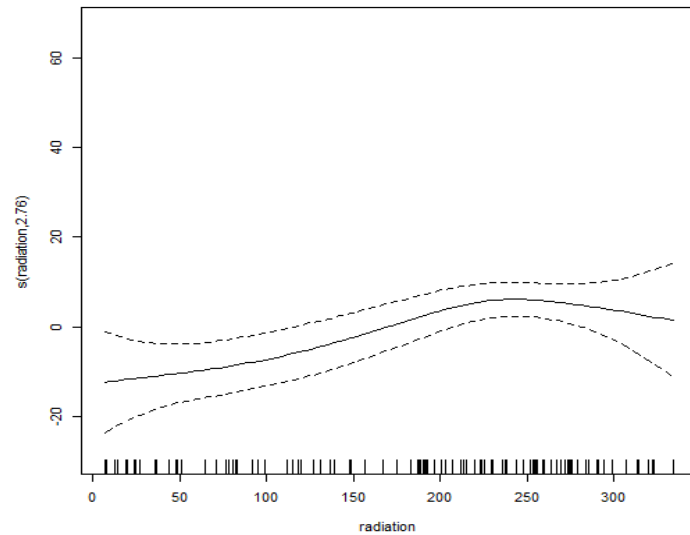
- Fit models with less assumptions
  - No 'nice' polynomial shapes are necessary
  - No variance assumptions

When Can I Fit a GAM?

- You can fit a GAM with any data where you might try fitting LMs, GLMs, and GLMMs
  - GAMs are more general and with less assumptions
- Common Examples
  - LDF fitting
  - Large data sets with complicated interaction effects
  - Models with many parameters but not a lot of data per parameter
  - Fitting a smoothed trend line that allows the trend to vary by year

GAM Tradeoffs

Advantages

1. Useful for non-parametric and semi-parametric data

2. Useful when data doesn't fit LM/GLM assumptions

3. Can paste splines directly into Excel

Disadvantages

1. Output may be more difficult to interpret to regulators and business side

2. Must be wary of over-fitting
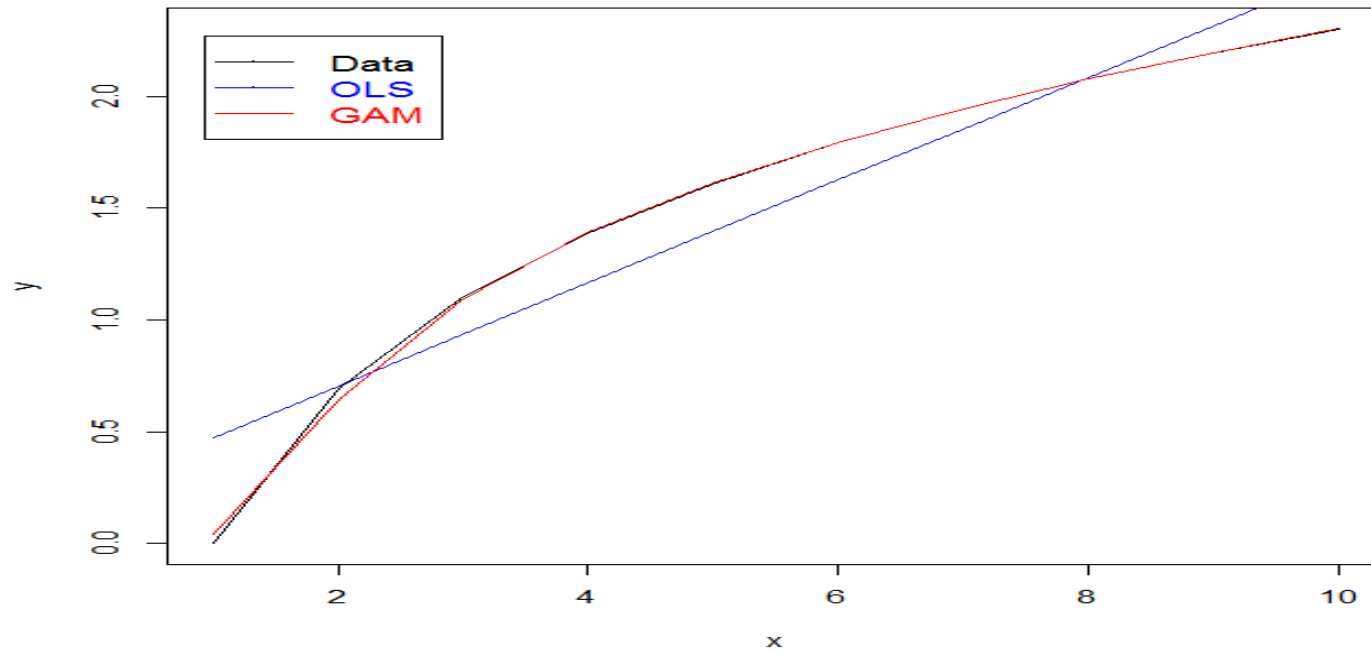
Let Software Do the Hard Work!

R

- Packages
  - gam
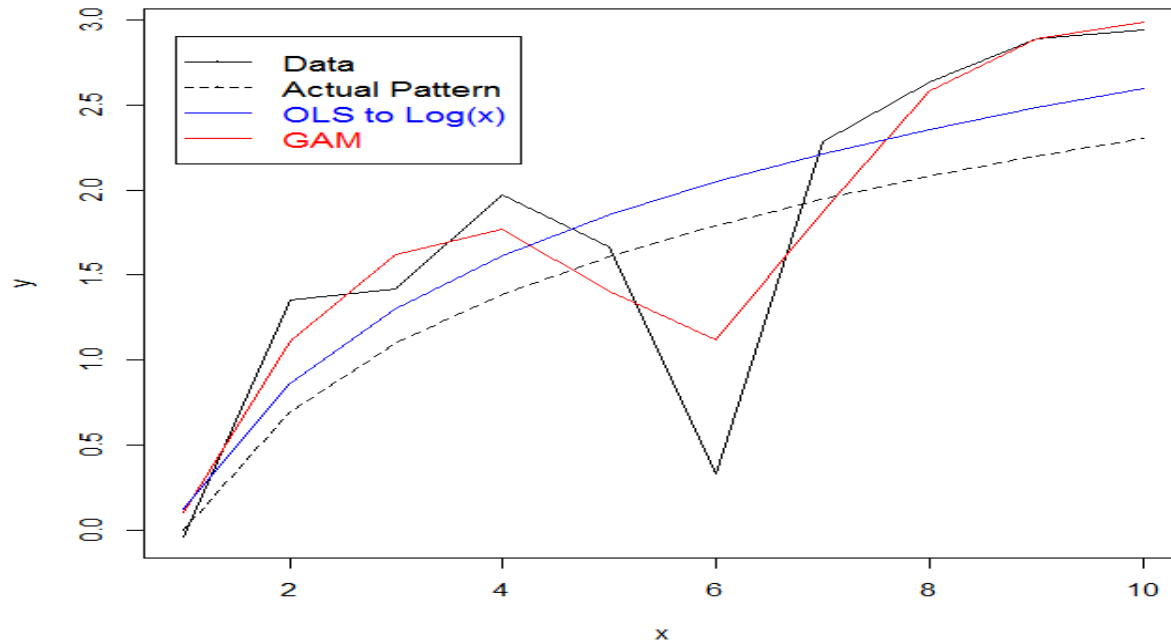 - mgcv – this package automatically selects smoothing factors

SAS

- PROC GAM
  - SAS 9.2

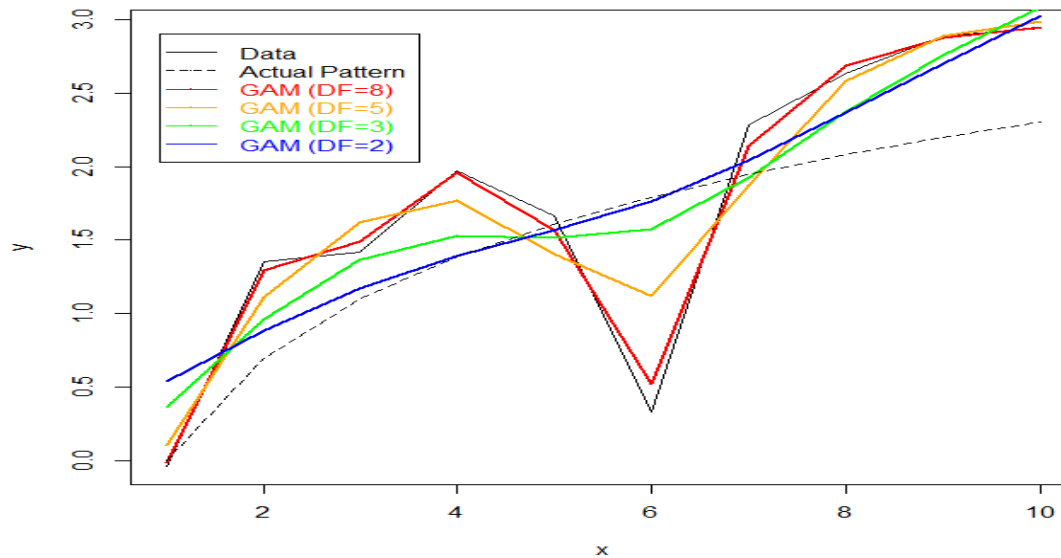# Simple GAM Example

▶ Smoothing to data can provide a very good fit

GAM Fitting to Noisy Data

- ▶ Smoothing to data can sometimes cause over-fitting though
- ▶ If a good parametric fit exists, use that instead

Amount of Smoothing Can be Varied

▶ DF = Degrees of Freedom = the number of parameters we are using to smooth

▶ A GAM ranges from a linear curve to fitting each point exactly

Some R Code for the Curious
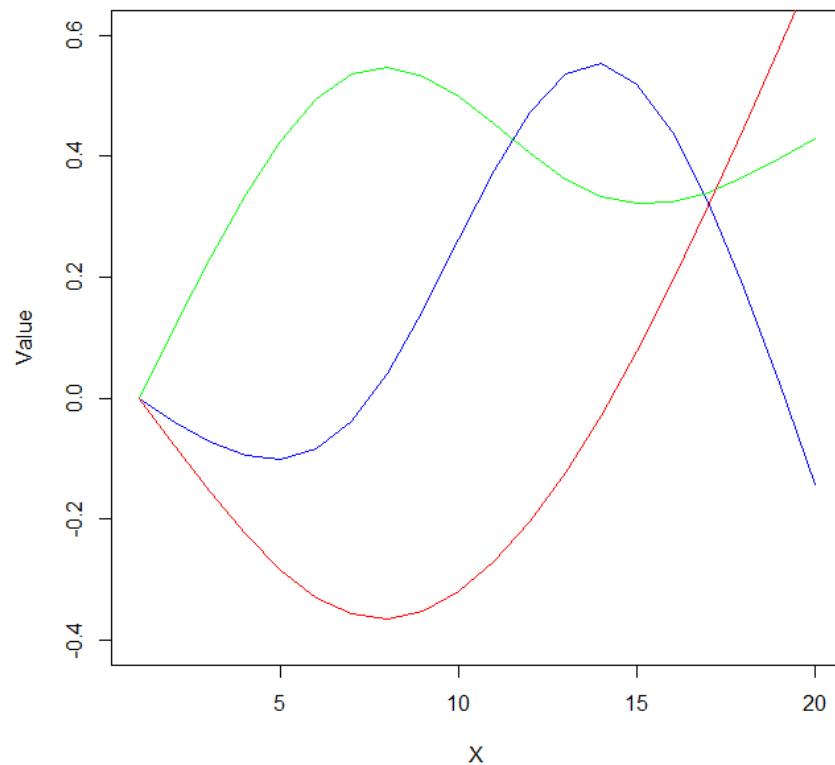
- x <- 1:10
- y <- log(x)

- plot(x, y, type='l')

- fit.lm <- lm(y ~ x)
- lines(predict(fit.lm), col='blue')

- library(gam)
- fit.gam <- gam(y ~ s(x, df=5))
- lines(predict(fit.gam), col='red')

Splines

- GAMs work by generating splines
- These can also be copied and pasted into Excel

- In R:
  - library(splines)
  - ns(1:20, df=3)

GAM Practical Application: LDF Fitting

- LDF patterns – Difficult to find a good parametric curve

- A GAM can be used to help smooth the curve to the data

- Will show an approach here that combines best features of two published models: the Inverse Power Curve (Extrapolating, Interpolating, and Smoothing Development Factors, Sherman, 1984) and England and Verrall's GAM model (A Flexible Framework For Stochastic Claims Reserving, 2001)

GAM Practical Application: LDF Fitting

▶ Inverse Power Curve

  ▶ Good: Simple procedure that can fit a portion of the LDFs well

  ▶ Bad: Struggles in many lines to provide a good fit to the entire curve

▶ England & Verrall's GAM Model

  ▶ Good: Uses a GAM to get a nice fit to the incremental loss pattern (within the common GLM loss development framework)

  ▶ Bad:

    ▶ Negative values difficult to deal with

    ▶ Some of the resulting LDFs can be difficult to interpret when comparing to the empirical LDFs

    ▶ More difficult to implement: Need to find correct Tweedie power

    ▶ Can't implement in Excel

    ▶ Hard to incorporate credibility (teaser)

Proposed Approach

- Smoothed Inverse Power Curve using GAMs (Korn 2015?)

- Inverse Power Curve:
  - $\log(LDF - 1) = A + B \log(t)$

- Smoothed Inverse Power Curve:
  - $\log(LDF - 1) = A + s(\log(t))$
    - Where "s" means GAM smoothing
    - (Note the smoothing is done on $\log(t)$)

Smoothed Inverse Power Curve
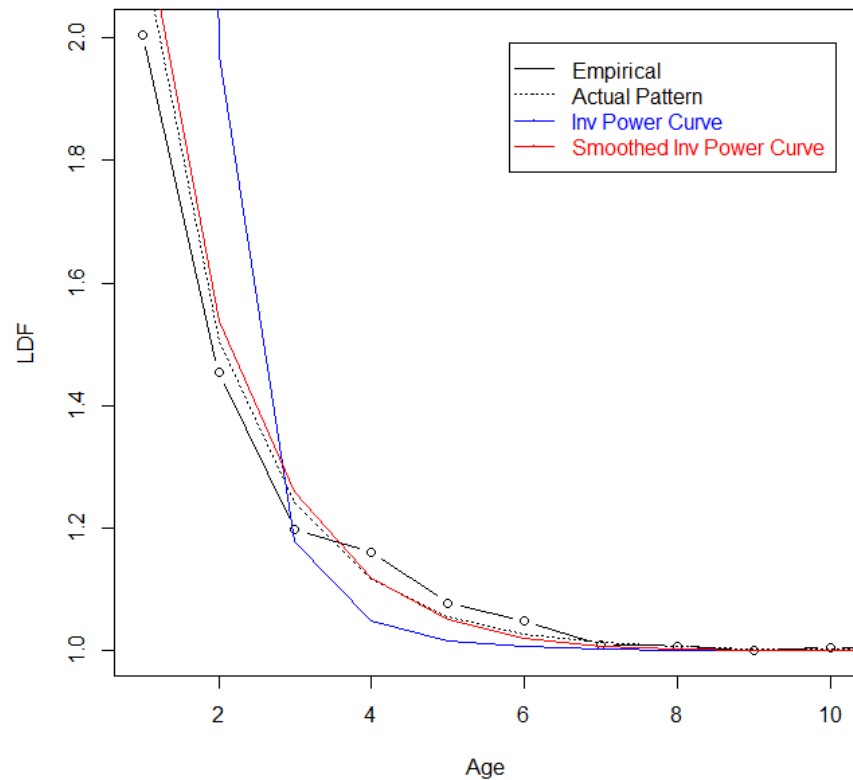
- ► Comparison of two approaches on simulated data
- ► This same pattern has been observed on actual data, where the inverse power curve has trouble making the "turn"
- ► The smoothed inverse power curve does a good job of smoothing out the volatility
- ► (No, I did not fish for a good example)

Still Interested in GAMs?

- <u>Elements of Statistical Learning</u>
  - By Hastie and Tibshirani
  - Free download:
    <u>http://web.stanford.edu/~hastie/local.ftp/Springer/OLD/ESLII_print4.pdf</u>

- <u>Stochastic Claims Reserving in General Insurance</u>
  - By England and Verrall

Review of Linear Models (Again)

Classical Linear Model

- Response: $Y \sim N(X\beta, \sigma^2)$
- $X\beta$ is a linear function that describes how the expected values vary based on characteristics in the data
- Linear: $\beta_0 + \beta_1 X_1^2 + sin(\beta_2 X_2)$
- Non-linear: $\beta_1 X_1 e^{\beta_2 X_2}$
- Constant Variance

Generalized Linear Model

- Response: Poisson, Gamma, Binomial, etc.
- $Y \sim F(\pi, R)$
- Expected Value:
  $G(E[Y])^{-1} = X\beta$
- Variance is a function of expected value
- Responses are independent

Correlated Losses

- But in the real world losses may not be independent

  - Why?

- **Hierarchical Data** – Correlation can exist among loss data when the risks come from the same territory or region
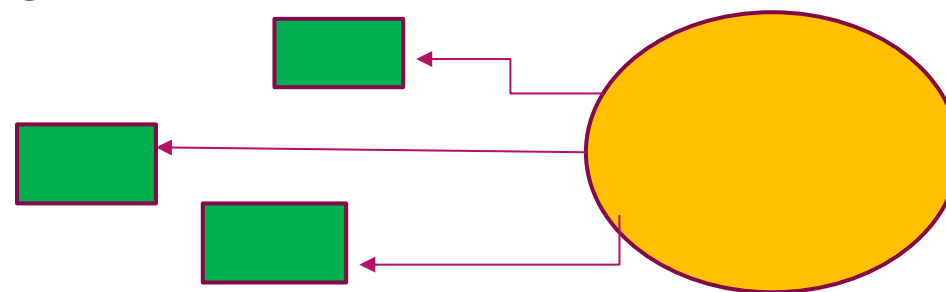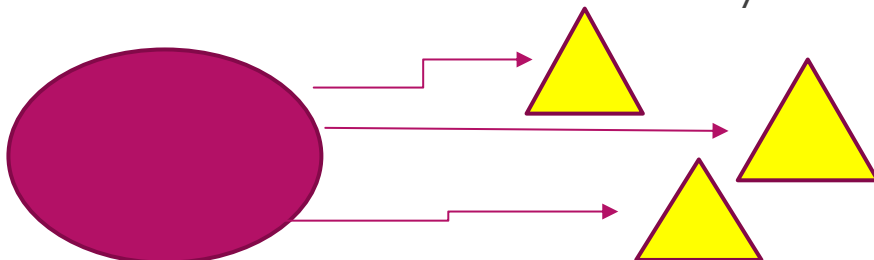
- **Repeated Measures** - Unless you have 0% retention, correlation can exist among records as some of them will represent the same risk repeatedly observed over time

Mixed Effect Models to the Rescue!

▶ Linear Predictor contains fixed effects and random effects

▶ $X\beta + Zb$

▶ $Z \sim N(0, G)$

    ▶ G is a covariance matrix that can reflect the extra variability and the correlation within the levels of a territory or across time

    ▶ Flexible enough to specify different G side covariance structures

▶ Response can still be Normal or from Exponential Class

LM	LMM
GLM	GLMM

What if I Use a LM/GLM Anyway?

Linear Model

▶ LM thinks it estimates this:
$$Y \sim N(X\beta, R)$$

▶ But it actually estimates this:
$$Y \sim N(X\beta, G + R)$$

▶ **Result:** correct parameter estimates but incorrect covariance estimate and distorted alpha levels

Generalized Linear Model

▶ GLM thinks it estimates this:
$$Y \sim F(\pi, R)$$

▶ But it actually estimates this:
$$Y \sim F(\tilde{\pi}, TGT + R)$$

▶ **Result:** Incorrect parameter estimates and incorrect covariance structure and distorted alpha levels

Don't Do Heavy Math by Yourself! Use Software!

R

- Package: lme4
  - Can fit common distributions but **not** the over-dispersed Poisson or Tweedie

SAS

- PROC GLIMMIX
  - SAS 9.1 and later
  - Can fit common distributions **and** over-dispersed Poisson
  - Uncertain about Tweedie

Simulation Example - GLMM

- Random effects:
- Groups   Name   Variance Std.Dev. Corr
- territory (Intercept)   0.1525   0.3905
-                   A1   0.3756   0.6129   0.27
- Number of obs: 20000, groups: territory, 100

- Fixed effects:
-              Estimate Std. Error z value Pr(>|z|)
- (Intercept) 2.03836   0.03917   52.03   <2e-16 ***
- A1   0.61291   0.06135   9.99   <2e-16 ***

Simulation Example - GLM

- Coefficients:

-                Estimate Std. Error   z value Pr(>|z|)

- (Intercept) 2.303576   0.002483   927.8   <2e-16 ***

- A1   0.671555   0.002483   270.5   <2e-16 ***

Expected Counts

	Actual	GLMM	GLM
Base Class	7.4	7.6	10.0
A1	12.2	14.1	19.6

Practical Application – Credibility Weighting

- A GLMM with a normal distribution and an identity link will produce identical results as the Buhlmann-Straub method

  - Benefits of GLMM:

    - Easier to automate – no need to manually calculate the within and between variances

    - More flexibility

      - More complicated regression models, such as hierarchal and multi-dimensional

      - Ability to handle different link functions (e.g. log, logit), non-normal errors, and continuous variables

  - A disadvantage is that a GLMM is harder to use from Excel

Credibility Weighting of Expected Loss Ratios/Costs

- Don't just credibility weight the IBNR portion! – This will be credibility weighting only half of the data.

- Don't credibility weight the selected ultimates from a BF (or similar) method! – That would be including what did NOT happen (and lowering the variance = too much credibility).

- Use the observed experience!

Coin Flipping Analogy

▶ First Time: 20 Flips

   ▶ 15 Heads, 5 Tails = 75% Heads

▶ Second Time (Same Coin): 5 Flips (out of 20)

   ▶ 0 Heads, 5 Tails

   ▶ IBNR (BF Method): 11.25 Heads, 3.75 Tails

   ▶ Ultimate: 11.25 Heads, 8.75 Tails = 56.25% Heads

▶ What's the variance??

Credibility Weighting of Expected Loss Ratios/Costs

▶ Instead, use a Cape-Cod-like method:

▶ LR per Year = Reported Losses / Used Premium (= chain ladder)

▶ Initial Weight per Year = Used Premium

▶ Then apply an off-balance factor so that the total weight for each segment equals the actual premium

Credibility Weighting of Expected Loss Ratios/Costs

- Weights:
  - GLMMs use the same weight for credibility as they do for the regression
  - For calculating the variances, the weight is assumed to be the number of observations
    - Premium Weights = Full Credibility
    - Claim Count Weights = Inconsistencies

  - To reconcile: (Really use weights as above instead of Premium)
    - K = Claim Count / Premium (for all policies)
    - For each policy, Weight(i) = Premium(i) x K
    - Total weight will be consistent with true number of observations and we will still be weighting by premium

Credibility Weighting of Expected Loss Ratios/Costs

- Structure of the GLMM:

  - Normal/Normal (is not the same as assuming that loss ratios are normally distributed)

  - Link function:

    - Log link: Multiplicative (dealing with 0s)
    - Identity link: Additive

Some R Code for the Curious

- library(lme4)
- fit <- glmer(lr ~ (1 | sic1) + (1 | sic2), weights=w, data=mydata, family=gaussian(link='log'))
- fixef(fit)
- ranef(fit)

Another Practical Use – Credibility Weighted Interaction Terms

- State x Industry Example:

  - No interaction: If New York is running 20% worse overall, it will be 20% worse in every single industry

    - Not Enough Information!

  - With interaction term: Every single state x industry combination will be assigned a loss ratio based on its experience alone

    - Not Enough Data!

  - Credibility weighted interaction term: If New York is running 20% worse overall, this will be the complement of credibility for each industry

    - Makes the most out of limited data

    - In R: "(1 | state:industry)"

Uneven Hierarchies
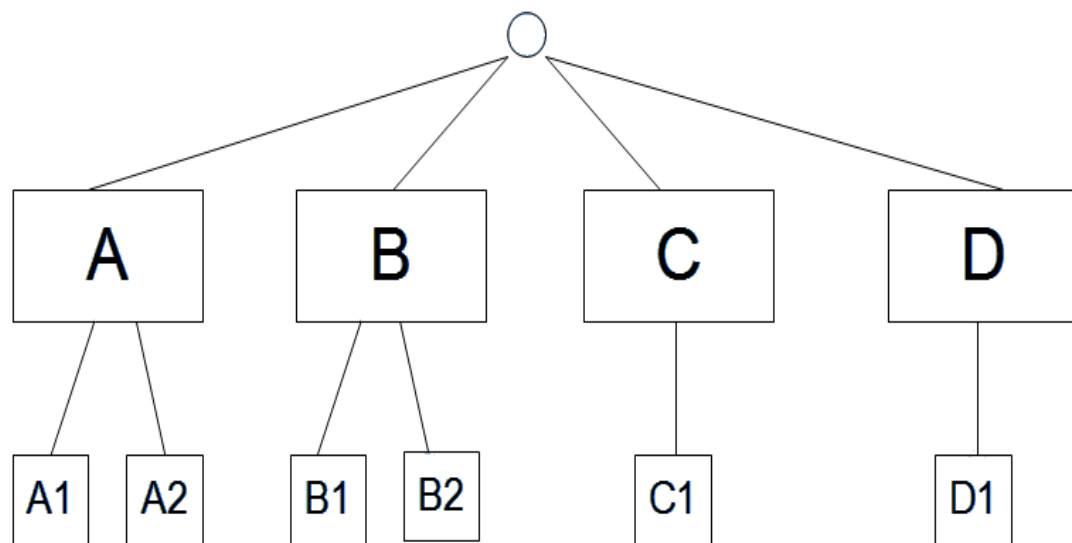
- A GLM will not produce coefficient values for C1 and D1 since they are redundant

- A GLMM will → Double Credibility!

- To handle, create a dummy variable that is 1 for A & B cells, and 0 for C & D cells

- For the lowest subgroup, create the random effect as a slope parameter on this dummy variable

- This will cause C1 and D1 to not be assigned coefficients

- In R: (1 | group) + (dummy | subgroup)

Other Practical Uses

- Incorporating credibility into pricing or other GLMs

- Credibility weighting of trend (if you have enough data)

Still Interested in GLMMs?

▶ Further Reading:

▶ Generalized Linear Mixed Models: Modern Concepts, Methods and Applications

  ▶ By Walter Stroup

  ▶ Examples of SAS

▶ Mixed-Effect Models in S and S-Plus

  ▶ By Pinheiro and Bates

  ▶ Written for R