EMB America

GLM II: Basic Modeling Strategy

CAS Predictive Modeling Seminar

Geoff Werner
September 2005

* Background
* Modeling Steps
  - Get Data
  - Initial Selections
  - Test Error/Link
  - Prelim Interrogation
  - Build Model
  - Validate Final Model
  - Combine Models
* Business Restrictions
* Summary

Basic Modeling Session

PURPOSE: To discuss basic modeling strategies and techniques for building appropriate GLM models

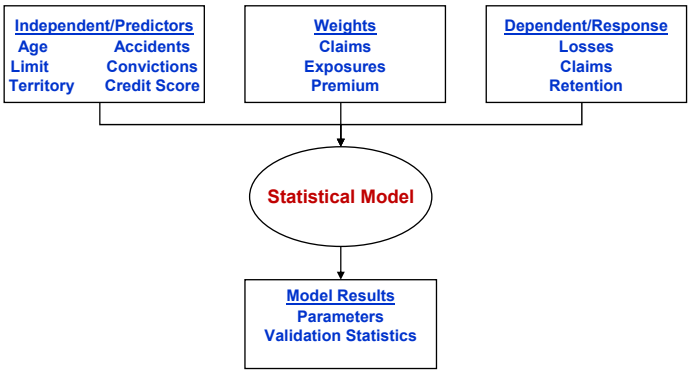
OUTLINE

- Background
- Basic Predictive Modeling Steps
  1. Get clean data
  2. Select an initial error structure, link function, and model structure
  3. Test error structure/link function
  4. Preliminary Interrogation
  5. Build predictive models iteratively
  6. Validate final predictive model
  7. Combine models, if modeling frequency and severity
- Implement Business Restrictions
- Summary

• Background
• Modeling Steps
- Get Data
- Initial Selections
- Test Error/Link
- Prelim Interrogation
- Build Model
- Validate Final Model
- Combine Models
• Business Restrictions
• Summary

Purpose of Predictive Modeling

- To predict a response variable using a series of explanatory variables (or rating factors)



Same techniques apply regardless of what is being modeled

EMB America

• Background
• Modeling Steps
- Get Data
- Initial Selections
- Test Error/Link
- Prelim Interrogation
- Build Model
- Validate Final Model
- Combine Models
• Business Restrictions
• Summary

Generalized Linear Models (GLMs)

- GLM is a multivariate method and considers all factors simultaneously

Response Variable = Systematic Component + Random Component

Systematic Component: Signal

Random Component: Noise

$y = h(\text{Linear Combination of Rating Factors}) + \text{Error}$

$g=h^{-1}$ is called the LINK function

Combination of rating factors is the model structure

Error structure reflects the underlying process

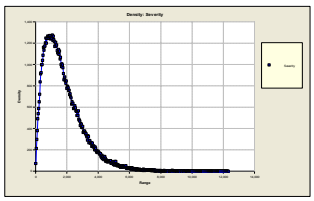
EMB America

GLM Building Blocks: Error Structure

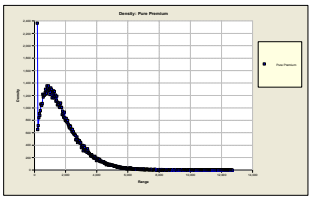
• Background
• Modeling Steps
- Get Data
- Initial Selections
- Test Error/Link
- Prelim Interrogation
- Build Model
- Validate Final Model
- Combine Models
• Business Restrictions
• Summary

y = h(Linear Combination of Rating Factors) + **Error**

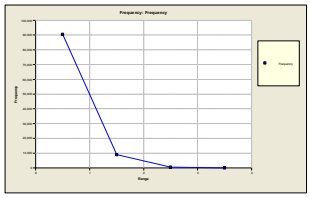
- Reflects the variability of the underlying process and can be <u>any</u> distribution within the exponential family, for example:



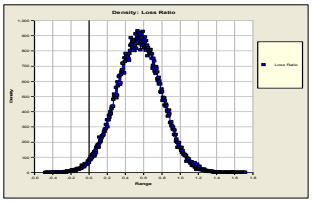
- Gamma consistent with severity modeling, may want to try Inverse Gaussian



- Tweedie consistent with pure premium modeling



- Poisson consistent with frequency modeling



- Normal useful for a variety of applications

EMB America

GLM Building Blocks: Model Structure

• Background
• Modeling Steps
- Get Data
- Initial Selections
- Test Error/Link
- Prelim Interrogation
- Build Model
- Validate Final Model
- Combine Models
• Business Restrictions
• Summary

y = h(**Linear Combination of Rating Factors**) + Error

- Include variables that are predictive, exclude those that are not
  - Gender may not have major impact on theft severity
- Simplify some rating factors, if full inclusion not necessary
  - Some levels within a particular predictor may be grouped together (e.g., 50-54 year olds)
  - A curve may replicate the signal (e.g., amount of insurance)
- Complicate model if the relationship between levels of one variable depends on another characteristic
  - The difference between males and females depends on age

EMB America

• Background
• Modeling Steps
- Get Data
- Initial Selections
- Test Error/Link
- Prelim Interrogation
- Build Model
- Validate Final Model
- Combine Models
• Business Restrictions
• Summary

GLM Building Blocks: Link Functions

y = **h(**Linear Combination of Rating Factors**)** + Error

- Link function ($g=h^{-1}$) chosen to based on how the factors are related to produce the best signal:
  - Log: variables related multiplicatively (e.g., risk modeling)
  - Identity: variables related additively (e.g., risk modeling)
  - Logit: retention or risk modeling
  - Reciprocal: canonical link for gamma distribution
  - Mixed: additive/multiplicative rating algorithms

• Background
• Modeling Steps
- Get Data
- Initial Selections
- Test Error/Link
- Prelim Interrogation
- Build Model
- Validate Final Model
- Combine Models
• Business Restrictions
• Summary

Example: Log Link

- The signal allows us to populate rating tables:

Premium = Base Premium x Policyholder Age x Rating Area

h(linear combination of rating factors)

Base Premium = 1,000 **=exp(6.908)**

Policyholder Age (p)	Relativity	
Youthful	1.700	=exp(.531)
Adult	1.000	=exp(0.000)
Mature	0.800	=exp(-0.223)
Seniors	1.100	=exp(0.095)

Rating Area (r)	Relativity	
A	0.900	=exp(-0.105)
B	1.000	=exp(0.000)
C	1.150	=exp(0.140)
D	1.300	=exp(0.262)
E	1.500	=exp(0.405)

- The rating structure is multiplicative and the premium for a youthful policyholder living in Area C is:

$1,955=$1,000 x 1.700 x 1.150

= exp(6.908) * exp(0.531) * exp(0.140)

= exp(6.908 + 0.531 + 0.140)

= exp(b+p+r)

* Background
• Modeling Steps
- Get Data
- Initial Selections
- Test Error/Link
- Prelim Interrogation
- Build Model
- Validate Final Model
- Combine Models
* Business Restrictions
* Summary

Basic Predictive Modeling Steps

1. Get clean data
2. Select initial error structure, link function and model structure
3. Test error structure/link function
4. Preliminary investigation
5. Build model structure
   - Include/exclude factors
   - Group rating levels
   - Fit curves
   - Include interactions
6. Validate final predictive model
7. Combine models

* Background
• Modeling Steps
- Get Data
- Initial Selections
- Test Error/Link
- Prelim Interrogation
- Build Model
- Validate Final Model
- Combine Models
* Business Restrictions
* Summary

Get Clean Data

- Data remains the number 1 issues for companies building predictive models
- Issues to be resolved
  - Null records or bad data (e.g., 10 year old drivers), especially for variables not used in rating
  - Poor linkage between losses and policy characteristics
  - Too much pre-banding of data
  - No mapping of old groupings into new groupings (e.g., boundary changes)
  - For auto, no linkage between operator, vehicle, and policy characteristics
  - Inconsistency between variables (e.g., 30 year olds living in a retirement community)
- Some issues cannot be resolved, impact on analysis depends on the type and extent of the problem

* Background
• Modeling Steps
- Get Data
- Initial Selections
- Test Error/Link
- Prelim Interrogation
- Build Model
- Validate Final Model
- Combine Models
* Business Restrictions
* Summary

Select Error Structure/Link Function

Observed Response	Most Appropriate Link Function	Most Appropriate Error Structure	Variance Function
--	--	Normal	μ^0
Claim Frequency	Log	Poisson	μ
Claim Severity	Log	Gamma	μ^2
Claim Severity	Log	Inverse Gaussian	μ^3
Risk Premium	Log	Gamma or Tweedie	μ^T
Retention Rate	Logit	Binomial	$\mu (1-\mu)$
Conversion Rate	Logit	Binomial	$\mu(1-\mu)$

* Background
• Modeling Steps
- Get Data
- Initial Selections
- Test Error/Link
- Prelim Interrogation
- Build Model
- Validate Final Model
- Combine Models
* Business Restrictions
* Summary

Select "Initial" Model

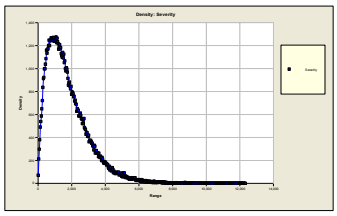
- Modeling is an iterative process, but you must select an initial point
- Reasonable Options
  - All variables
  - Prior models
  - Model for a similar cause of loss
  - All known important variables (e.g., rating factors)
  - Stepwise regressions (i.e., forward and backward)
- Oftentimes initial model is a "simple" model
  - May require minimal simplifications to achieve initial fit
  - May include known interactions

GLM II:
Basic Modeling

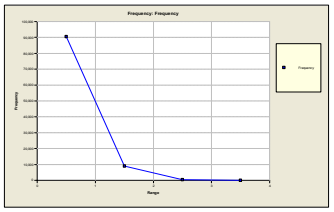
- Background
- Modeling Steps
  - Get Data
  - Initial Selections
  - Test Error/Link
  - Prelim Interrogation
  - Build Model
  - Validate Final Model
  - Combine Models
- Business Restrictions
- Summary

Test Error Structure/Link Function
Distribution Analysis

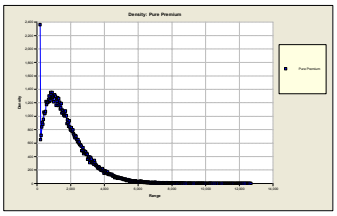
- Examine plots of the data (e.g., size of loss distribution)



- Consistent with a gamma distribution



- Consistent with a Poisson distribution



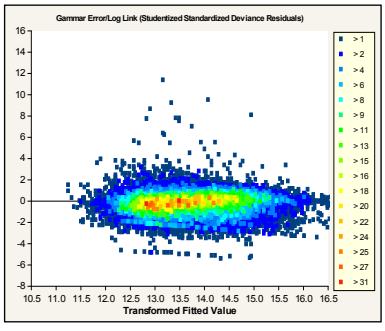
- Consistent with aTweedie distribution

GLM II:
Basic Modeling

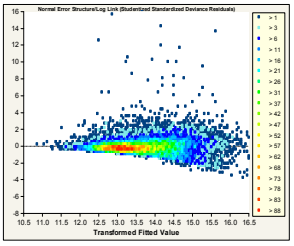
- Background
- Modeling Steps
  - Get Data
  - Initial Selections
  - Test Error/Link
  - Prelim Interrogation
  - Build Model
  - Validate Final Model
  - Combine Models
- Business Restrictions
- Summary

Test Error Structure/Link Function
Macro Residual Analysis

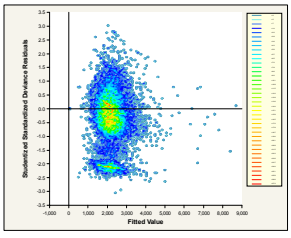
- Plot of <u>all</u> residuals tests selected error structure/link function



- Elliptical pattern is ideal



- Asymmetrical appearance suggests power of variance function is too low

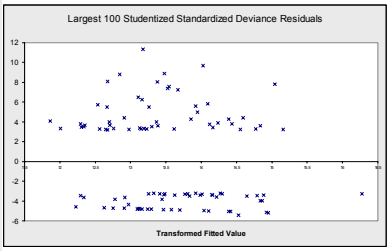


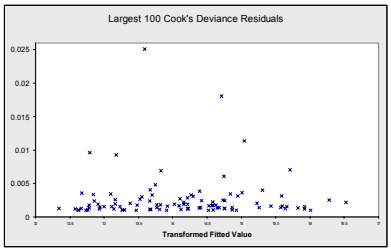
- Two concentrations suggests two perils: split or use joint modeling

- Background
- Modeling Steps
  - Get Data
  - Initial Selections
  - Test Error/Link
  - Prelim Interrogation
  - Build Model
  - Validate Final Model
  - Combine Models
- Business Restrictions
- Summary

Test Error Structure/Link Function
Micro Residual Analysis

- Examine largest residuals…





- **Standardized deviance gives a measure of "fit"**
- **Cook's deviance gives a measure of "influence"**

Influence

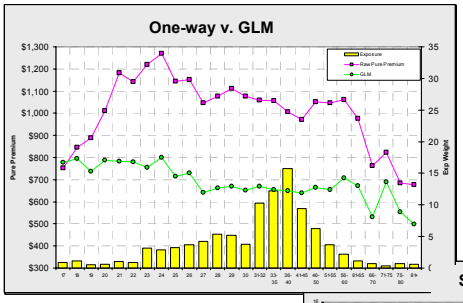
Fit	Small	Large
Good	OK	OK
Poor	OK	Problem

- "Problem" points may require further investigation

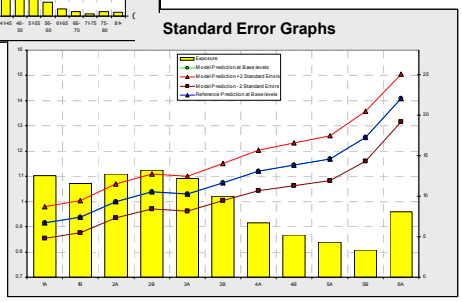
- Background
- Modeling Steps
  - Get Data
  - Initial Selections
  - Test Error/Link
  - Prelim Interrogation
  - Build Model
  - Validate Final Model
  - Combine Models
- Business Restrictions
- Summary

Preliminary Investigation

- Traditional statistics and simple graphs provide "quick" feel



- Highlights what others within your company "know"

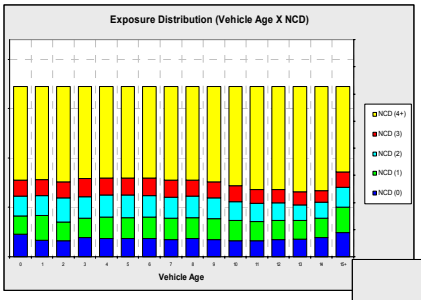


- Quickly highlight trends in your data

Preliminary Investigation

- Background
- Modeling Steps
  - Get Data
  - Initial Selections
  - Test Error/Link
  - Prelim Interrogation
  - Build Model
  - Validate Final Model
  - Combine Models
- Business Restrictions
- Summary

- Statistics can (e.g., Cramer's V) identify correlated variables

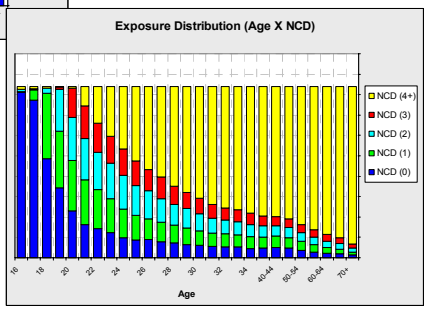


Low Correlation (.025)

- Distribution of number of years claims free about the same for each vehicle age

High Correlation (.253)

- Older drivers are more likely to be claim-free

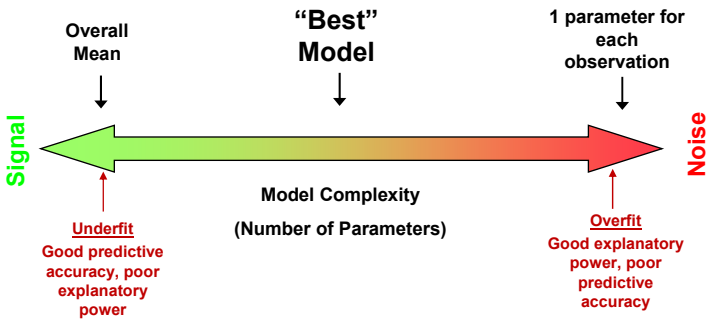


EMB America

Building the "Best" Model

- Background
- Modeling Steps
  - Get Data
  - Initial Selections
  - Test Error/Link
  - Prelim Interrogation
  - Build Model
  - Validate Final Model
  - Combine Models
- Business Restrictions
- Summary

- To produce a sensible model that explains recent historical experience and is likely to be predictive of future experience.



EMB America

Simple Model Parameter Notation

* Background
• Modeling Steps
- Get Data
- Initial Selections
- Test Error/Link
- Prelim Interrogation
- Build Model
- Validate Final Model
- Combine Models
* Business Restrictions
* Summary

- Example: 2 rating variables (Age and Gender)

Simple Model: Age + Gender

	Male (Base)	Female
16	$\beta_0 + \beta_{16}$	$\beta_0 + \beta_{16} + \beta_F$
17	$\beta_0 + \beta_{17}$	$\beta_0 + \beta_{17} + \beta_F$
:	:	:
30 (Base)	β_0	$\beta_0 + \beta_F$
31	$\beta_0 + \beta_{31}$	$\beta_0 + \beta_{31} + \beta_F$
:	:	:
64	$\beta_0 + \beta_{64}$	$\beta_0 + \beta_{64} + \beta_F$
65+	$\beta_0 + \beta_{65+}$	$\beta_0 + \beta_{65+} + \beta_F$

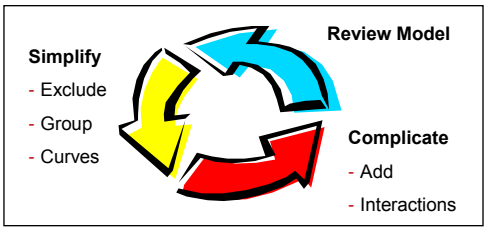
- Log link: $\text{Relativity}_{16,F} = \exp(\beta_0 + \beta_{16} + \beta_F)/\exp(\beta_0)$
- Identity link: $\text{Additive}_{16,F} = (\beta_0 + \beta_{16} + \beta_F) - (\beta_0)$

EMB America

Iterative Modeling

* Background
• Modeling Steps
- Get Data
- Initial Selections
- Test Error/Link
- Prelim Interrogation
- Build Model
- Validate Final Model
- Combine Models
* Business Restrictions
* Summary

- Modeling is an iterative process



How does the analyst decide the "Best" Model?

1. Examine parameters and standard errors to discern pattern and the "confidence" interval
2. Check consistency over time to see if trends are stable
3. Apply statistical tests (e.g., Chi-Squared) that compare models
4. Use judgement

EMB America

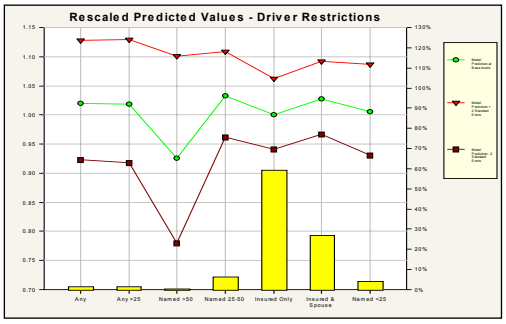
Modeling-Including/Excluding Factors

- Parameters/standard errors tell importance of factors and "confidence" in estimates

- Background
- Modeling Steps
  - Get Data
  - Initial Selections
  - Test Error/Link
  - Prelim Interrogation
  - Build Model
  - Validate Final Model
  - Combine Models
- Business Restrictions
- Summary

- If all the parameters are essentially the same or have very large standard errors, the factor may not be important

Name	Value	Standard Error	Standard Error (%)	Exp(Value)
Any	0.0174	0.04183	240.8	1.0175
Any>25	0.0212	0.04349	205.4	1.0214
Named >50	-0.0961	0.08120	84.5	0.9084
Named 25-50	0.0357	0.02194	61.4	1.0364
Insured Only				
Insured & Spouse	0.0255	0.01272	49.8	1.0259
Named <25	-0.0446	0.02663	59.7	0.9564



- Graph of parameters/standard errors and "horizontal line test" identifies importance of a factor

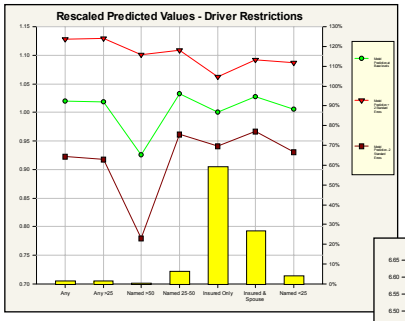
EMB America

Modeling-Including/Excluding Factors

- Look at patterns over time to identify true patterns versus anomalies

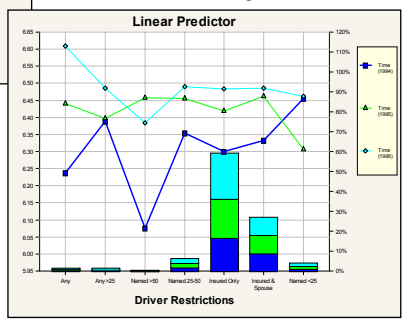
- Background
- Modeling Steps
  - Get Data
  - Initial Selections
  - Test Error/Link
  - Prelim Interrogation
  - Build Model
  - Validate Final Model
  - Combine Models
- Business Restrictions
- Summary

Parameter/Standard Errors



- Main effects graph may indicate a questionable estimate

Time Testing



- By testing the pattern over time can see if the same thing happens each year

EMB America

* Background
• Modeling Steps
- Get Data
- Initial Selections
- Test Error/Link
- Prelim Interrogation
- Build Model
- Validate Final Model
- Combine Models
* Business Restrictions
* Summary

Modeling-Including/Excluding Factors

- Goodness of fit tests (e.g., Chi-Squared) can be used to determine the explanatory power of a variable
  - Null hypothesis is that the models with and without the factor are the same

Chi-Squared

Model	With	Without
Deviance	8,906.4414	8,909.6226
Degrees of Freedom	18,469	18,475
Scale Parameter	0.4822	0.4823
Chi Square Test		78.6%

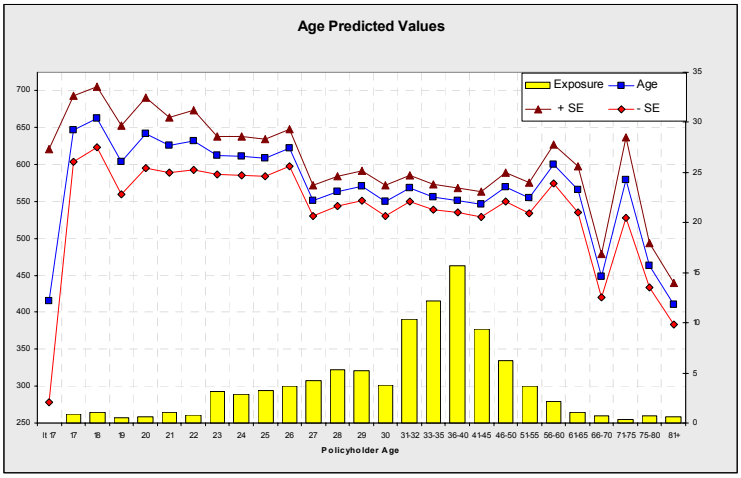
  - High score indicates null hypothesis should be accepted and the simpler model chosen → factor is not kept
  - Low score indicates null hypothesis should be rejected and more complicated model chosen → factor is kept

EMB America

* Background
• Modeling Steps
- Get Data
- Initial Selections
- Test Error/Link
- Prelim Interrogation
- Build Model
- Validate Final Model
- Combine Models
* Business Restrictions
* Summary

Modeling-Grouping Rating Levels

- Simplify trends in rating factors by grouping factor levels…



EMB America

- Background
- Modeling Steps
  - Get Data
  - Initial Selections
  - Test Error/Link
  - Prelim Interrogation
  - Build Model
  - Validate Final Model
  - Combine Models
- Business Restrictions
- Summary

Modeling-Grouping Rating Levels

- Examining the actual parameters and the weight of each level helps determine potential groupings

Name	Value	Standard Error	Standard Error (%)	Weight (%)	E(Value)
Lt 17	-0.2872	0.40047	139.4	0.0	0.7504
17	0.1597	0.06488	40.6	0.9	1.1731
18	0.1838	0.05642	30.7	1.1	1.2018
19	0.0915	0.07222	78.9	0.6	1.0958
20	0.1506	0.07009	46.6	0.6	1.1625
21	0.1254	0.05478	43.7	1.0	1.1336
22	0.1364	0.05916	43.4	0.8	1.1462
23	0.1038	0.03476	33.5	3.1	1.1094
24	0.1022	0.03559	34.8	2.9	1.1076
25	0.0979	0.03288	33.6	3.2	1.1029
26	0.1207	0.03098	25.7	3.7	1.1283
27	-0.0015	0.02947	1,929.7	4.2	0.9985
28	0.0221	0.02635	119.0	5.3	1.0224
29	0.0345	0.02611	75.7	5.2	1.0351
30	-0.0021	0.02925	1,396.1	3.8	0.9979
31-32	0.0291	0.02059	70.8	10.4	1.0295
33-35	0.0079	0.01941	244.6	12.2	1.0080
36-40				15.7	
41-45	-0.0103	0.02110	204.5	9.4	0.9897

EMB America

- Background
- Modeling Steps
  - Get Data
  - Initial Selections
  - Test Error/Link
  - Prelim Interrogation
  - Build Model
  - Validate Final Model
  - Combine Models
- Business Restrictions
- Summary

Modeling-Grouping Rating Levels

- Consider looking at the standard error of the parameter differences

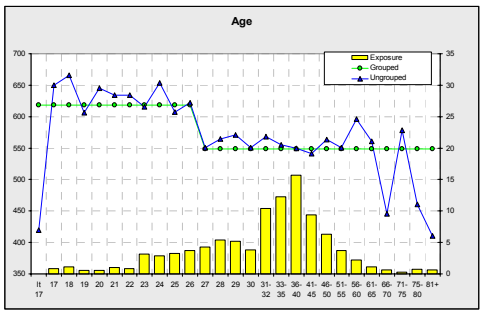
	Lt 17	17	18	19	20	21	22
Lt 17							
17	**90.4**						
18	**85.6**	**308.9**					
19	**107.2**	**132.7**	**91.2**				
20	**92.7**	**995.9**	**255.1**	**161.6**			
21	**97.8**	**236.1**	**127.0**	**254.7**	**332.7**		
22	**95.4**	**362.2**	**163.9**	**199.5**	**620.3**	**685.0**	
23	**102.6**	**124.2**	**76.9**	**618.2**	**158.1**	**273.1**	**193.0**
24	**103.1**	**122.4**	**76.6**	**719.3**	**154.6**	**259.0**	**186.9**
25	**104.2**	**112.5**	71.7	**1,182.8**	**140.8**	**217.5**	**165.4**
26	**98.4**	**176.5**	**96.1**	**258.8**	**246.0**	**1,250.8**	**399.8**
27	**140.4**	42.3	32.4	**80.8**	48.0	45.9	45.2
28	**129.6**	48.8	36.4	**106.9**	56.1	55.3	53.7
29	**124.6**	53.7	39.5	**130.3**	62.0	62.9	60.3
30	**140.7**	42.4	32.5	**80.6**	48.0	46.1	45.5
31-32	**126.6**	50.0	36.8	**116.4**	58.0	57.3	55.5
33-35	**135.7**	43.0	32.3	**86.7**	49.3	46.9	46.3
36-40	**139.4**	40.6	30.7	**78.9**	46.6	43.7	43.4

EMB America

Modeling-Grouping Rating Levels

- Simplify trends in rating factors by grouping factor levels…

* Background
• Modeling Steps
- Get Data
- Initial Selections
- Test Error/Link
- Prelim Interrogation
- Build Model
- Validate Final Model
- Combine Models
* Business Restrictions
* Summary



- Graph groupings to see if fit looks good visually

- Chi-squared test can compare the model with and without the grouping

Chi-Squared

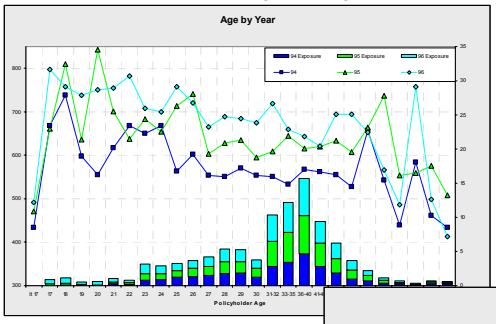
Model	No Group	Grouped
Deviance	8,906.4420	8,935.1260
Degrees of Freedom	18,469	18,494
Scale Parameter	0.4822	0.4831
Chi Square Test		27.7%

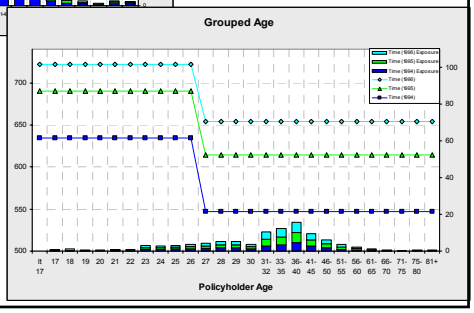
EMB America

Modeling-Grouping Rating Levels

- Explore if “indicated” groupings are consistent over time

* Background
• Modeling Steps
- Get Data
- Initial Selections
- Test Error/Link
- Prelim Interrogation
- Build Model
- Validate Final Model
- Combine Models
* Business Restrictions
* Summary





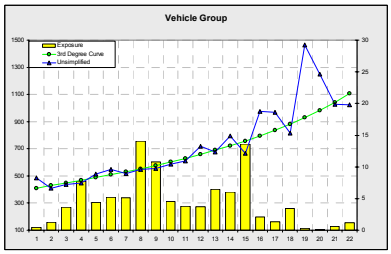
EMB America

Modeling-Fitting Curves (Variates)

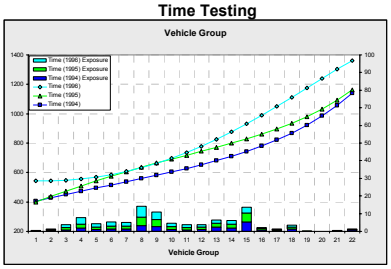
- Background
- Modeling Steps
  - Get Data
  - Initial Selections
  - Test Error/Link
  - Prelim Interrogation
  - Build Model
  - Validate Final Model
  - Combine Models
- Business Restrictions
- Summary

- Simplify trends in rating factors by fitting curves, splines, etc.

View Graphically



Time Testing



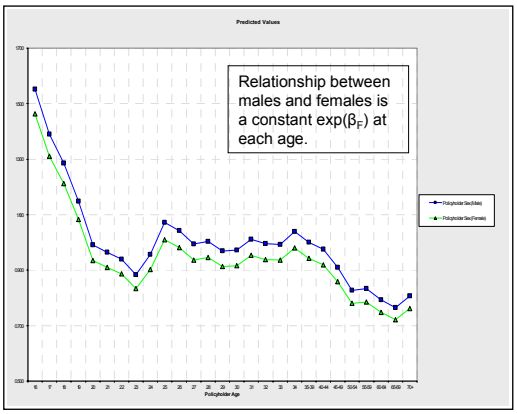
Chi-Squared

Model	No Curve	Curve
Deviance	8,906.4460	9,020.2270
Degrees of Freedom	18,469	18,487
Scale Parameter	0.4822	0.4879
Chi Square Test		0.0%

Modeling-Interactions
Graphical Display: Simple Model

- Background
- Modeling Steps
  - Get Data
  - Initial Selections
  - Test Error/Link
  - Prelim Interrogation
  - Build Model
  - Validate Final Model
  - Combine Models
- Business Restrictions
- Summary

Simple Model: Age + Gender



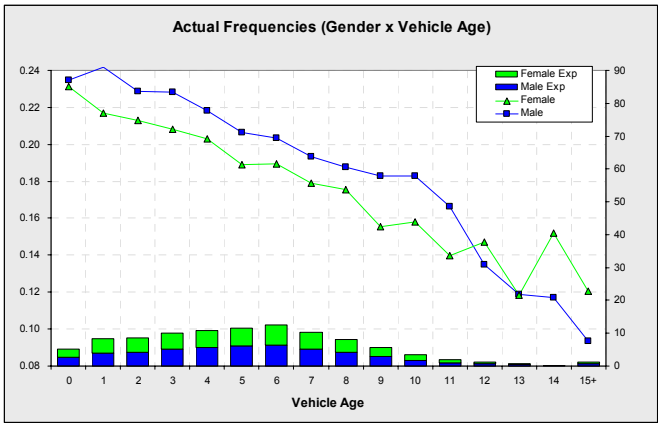
- Response correlation: the relationship between levels of one factor varies by the level of another factor
- Interactions are used to model response correlation

GLM II:
Basic Modeling

- Background
- **Modeling Steps**
  - Get Data
  - Initial Selections
  - Test Error/Link
  - Prelim Interrogation
  - **Build Model**
  - Validate Final Model
  - Combine Models
- Business Restrictions
- Summary

Modeling-Interactions
Low Response Correlation

- Compare model to actuals to see if "simple" model is adequate



  - Actual frequencies support relationship between male and female is basically constant for each vehicle age

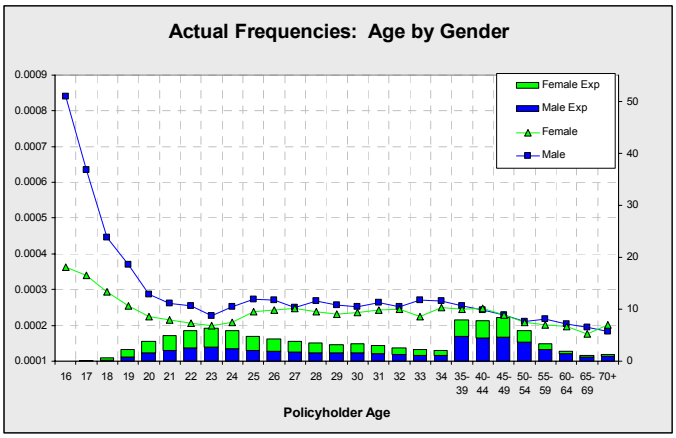
EMB America

GLM II:
Basic Modeling

- Background
- **Modeling Steps**
  - Get Data
  - Initial Selections
  - Test Error/Link
  - Prelim Interrogation
  - **Build Model**
  - Validate Final Model
  - Combine Models
- Business Restrictions
- Summary

Modeling-Interactions
High Response Correlation

- Compare model to actuals to see if "simple" model is adequate



  - Actual frequencies show relationship between male and female is very different for youthfuls and adults

EMB America

- Background
- Modeling Steps
  - Get Data
  - Initial Selections
  - Test Error/Link
  - Prelim Interrogation
  - Build Model
  - Validate Final Model
  - Combine Models
- Business Restrictions
- Summary

Modeling-Interactions
Parameter Notation: Full Interaction

- A full interaction allows the relationship between the levels of one variable to vary for each level of another variable

Model: Age+Gender

	Male (Base)	Female
16	$\beta_0+\beta_{16}$	$\beta_0+\beta_{16}+\beta_F$
17	$\beta_0+\beta_{17}$	$\beta_0+\beta_{17}+\beta_F$
:	:	:
30 (Base)	β_0	$\beta_0+\beta_F$
31	$\beta_0+\beta_{31}$	$\beta_0+\beta_{31}+\beta_F$
:	:	:
64	$\beta_0+\beta_{64}$	$\beta_0+\beta_{64}+\beta_F$
65+	$\beta_0+\beta_{65+}$	$\beta_0+\beta_{65+}+\beta_F$

Parameters: 51
- Mean: 1
- Age: 49
- Gender: 1
- AgexGender: 0

Model: Age+Gender+Age.Gender

	Male (Base)	Female
16	$\beta_0+\beta_{16}$	$\beta_0+\beta_{16}+\beta_F+\beta_{16,F}$
17	$\beta_0+\beta_{17}$	$\beta_0+\beta_{17}+\beta_F+\beta_{17,F}$
:	:	:
30 (Base)	β_0	$\beta_0+\beta_F$
31	$\beta_0+\beta_{31}$	$\beta_0+\beta_{31}+\beta_F+\beta_{31,F}$
:	:	:
64	$\beta_0+\beta_{64}$	$\beta_0+\beta_{64}+\beta_F+\beta_{64,F}$
65+	$\beta_0+\beta_{65+}$	$\beta_0+\beta_{65+}+\beta_F+\beta_{65+,F}$

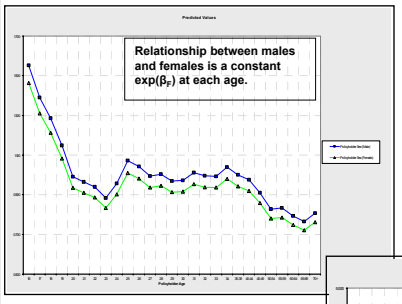
Parameters: 100
- Mean: 1
- Age: 49
- Gender: 1
- AgexGender: 49

EMB America

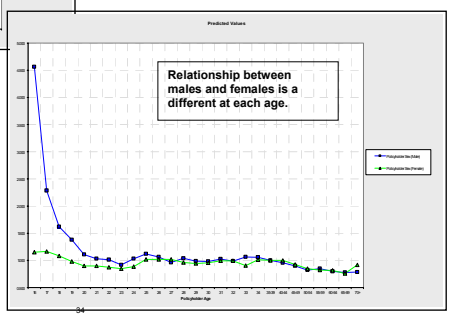
- Background
- Modeling Steps
  - Get Data
  - Initial Selections
  - Test Error/Link
  - Prelim Interrogation
  - Build Model
  - Validate Final Model
  - Combine Models
- Business Restrictions
- Summary

Modeling-Interactions
Testing Interactions

- View full interaction model



Simple Model: Age + Gender



Full Interaction Model:
Age + Gender + Age.Gender

EMB America

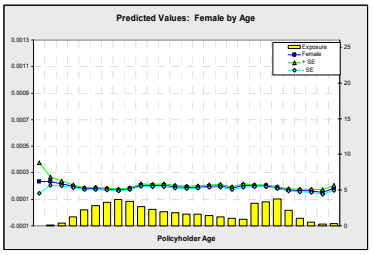
* Background
• Modeling Steps
- Get Data
- Initial Selections
- Test Error/Link
- Prelim Interrogation
- Build Model
- Validate Final Model
- Combine Models
* Business Restrictions
* Summary

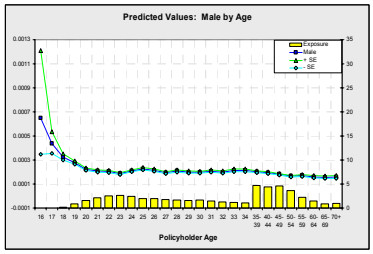
Modeling-Interactions
Testing Interactions

- View standard errors

Interaction Term	Value	Standard Error	Standard Error (%)	Weight
Female.16	-1.0235	0.78776	77.0	13,761
Female.17	-0.6174	0.24463	39.6	185,915
Female.18	-0.3981	0.11267	28.3	739,500
Female.19	-0.3382	0.07265	21.5	2,362,139
Female.20	-0.2112	0.06333	30.0	4,081,775
Female.21	-0.1384	0.05947	43.0	5,163,074
Female.22	-0.1467	0.05704	38.9	6,055,119
Female.23	-0.0782	0.05703	73.0	6,763,300
Female.24	-0.1536	0.05706	37.1	6,300,270
Female.25	-0.0972	0.05906	60.7	4,927,417
Female.26	-0.0431	0.06031	139.9	4,269,244
Female.27	0.0544	0.06364	116.9	3,672,472
Female.28	-0.0727	0.06477	89.1	3,438,810
Female.29	-0.0483	0.06761	140.0	2,970,306
Female.30	-0.0254	0.06693	263.3	3,027,278
Female.31	-0.0318	0.06849	215.1	2,724,535
Female.32	0.0033	0.07270	2,175.0	2,329,283
Female.33-35	-0.1597	0.07709	48.3	1,967,739
Female.36-39	-0.0376	0.07947	211.3	1,670,130
Female.40-44	0.0467	0.05185	111.1	6,166,191
Female.45-49	0.0297	0.05174	174.3	6,877,522
Female.50-54	0.0325	0.05973	183.8	3,957,251
Female.55-59	-0.0264	0.07412	281.0	1,998,839
Female.60-64	0.0228	0.09824	431.3	959,502
Female.65-69	-0.0168	0.13252	787.8	528,632
Female.70+	0.1593	0.12038	75.6	602,694

- In tabular format

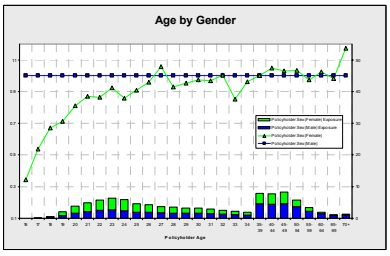




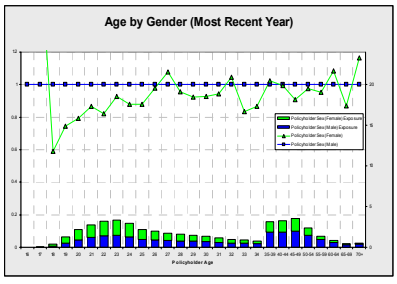
- Graphically

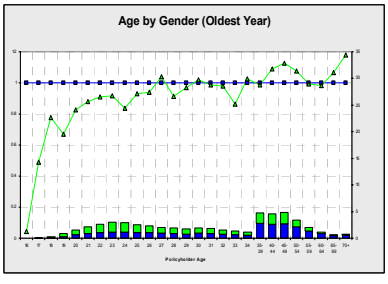
* Background
• Modeling Steps
- Get Data
- Initial Selections
- Test Error/Link
- Prelim Interrogation
- Build Model
- Validate Final Model
- Combine Models
* Business Restrictions
* Summary

Modeling-Interactions
Testing Interactions



- Examine full interaction model by year





* Background
• Modeling Steps
- Get Data
- Initial Selections
- Test Error/Link
- Prelim Interrogation
- Build Model
- Validate Final Model
- Combine Models
* Business Restrictions
* Summary

Modeling-Interactions
Testing Interactions

- Perform Type III test

Chi-Squared

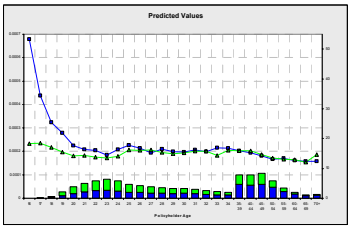
Model	Simple Model	W/ Interaction
Deviance	224,667.0000	224,771.0000
Degrees of Freedom	83	109
Scale Parameter	1.1615	1.1655
Chi Square Test		0.0%

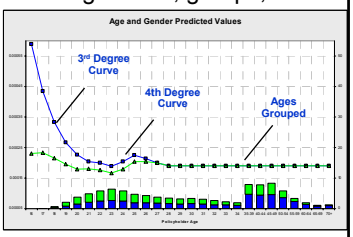
- Null hypothesis is that the models are the same
- Low score indicates null hypothesis should be rejected and more complicated model chosen → keep the interaction
- High score indicates null hypothesis should be accepted and the simpler model chosen → do not keep the interaction

* Background
• Modeling Steps
- Get Data
- Initial Selections
- Test Error/Link
- Prelim Interrogation
- Build Model
- Validate Final Model
- Combine Models
* Business Restrictions
* Summary

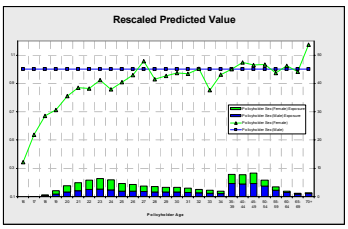
Modeling-Interactions
Simplifying Interactions

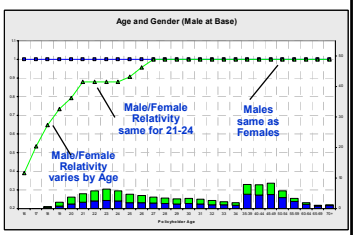
- Complex relationships can be simplified using curves, groups, etc.





- Simplify the age curve (i.e., the male age curve since male is base level)



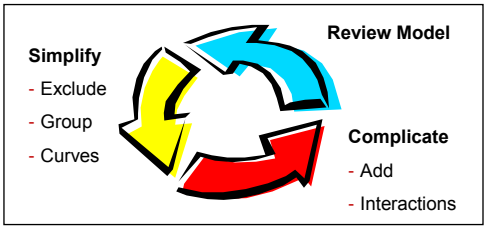


- Simplify the relationship between males and females

* Background
• Modeling Steps
- Get Data
- Initial Selections
- Test Error/Link
- Prelim Interrogation
- Build Model
- Validate Final Model
- Combine Models
* Business Restrictions
* Summary

Iterative Modeling

- Modeling is an iterative process

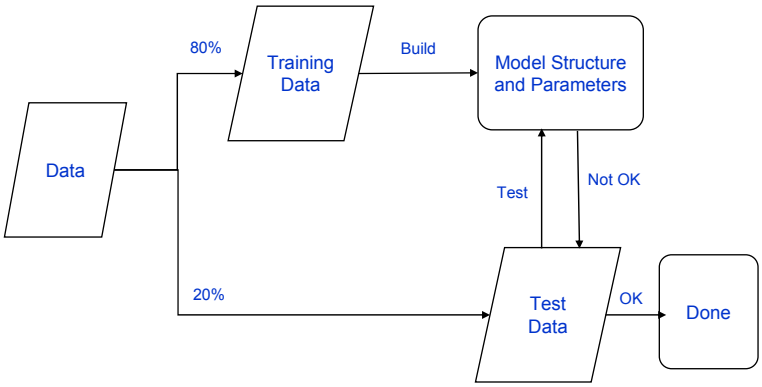


- May want to re-examine variables that were “in” or “out” before to make sure prior decisions are still appropriate!

* Background
• Modeling Steps
- Get Data
- Initial Selections
- Test Error/Link
- Prelim Interrogation
- Build Model
- Validate Final Model
- Combine Models
* Business Restrictions
* Summary

Review Model
Hold-Out Samples

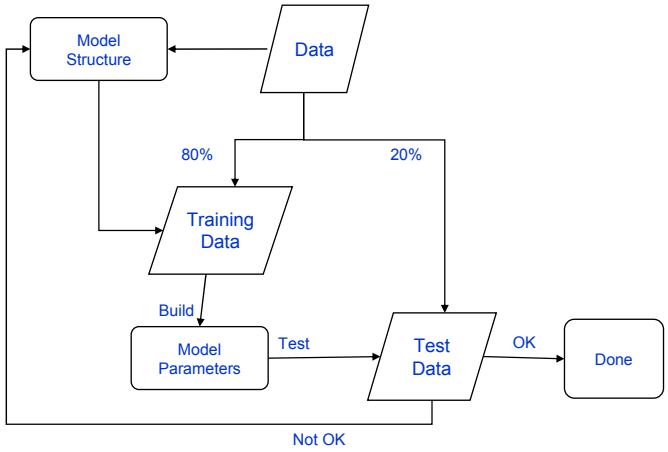
- If significant volume of data available, the modeler can perform full “Test/Training” approach



- Background
- Modeling Steps
  - Get Data
  - Initial Selections
  - Test Error/Link
  - Prelim Interrogation
  - Build Model
  - Validate Final Model
  - Combine Models
- Business Restrictions
- Summary

Review Model
Hold-Out Samples

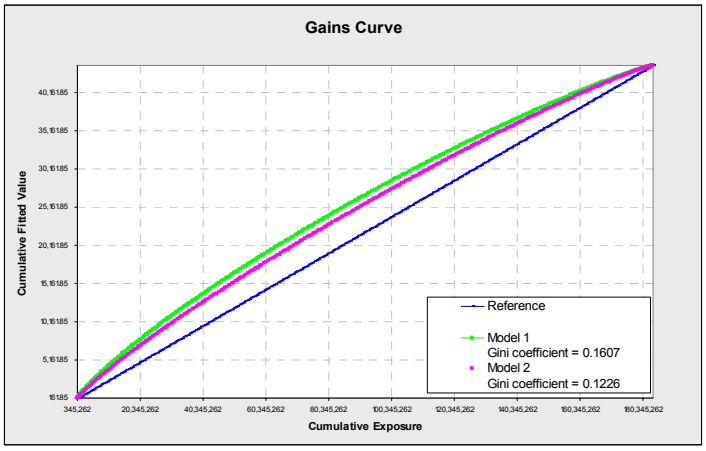
- Bootstrapping methods can be used when data volume prohibits full "Test/Training"



- Background
- Modeling Steps
  - Get Data
  - Initial Selections
  - Test Error/Link
  - Prelim Interrogation
  - Build Model
  - Validate Final Model
  - Combine Models
- Business Restrictions
- Summary

Review Model
Gains Curves

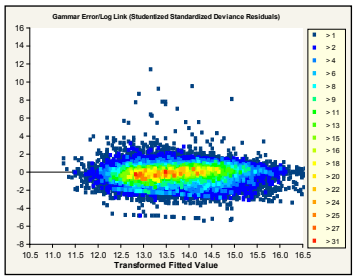
- Compares "lift" of models

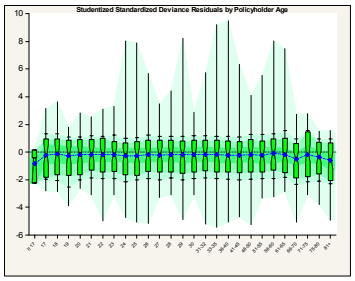


- Background
- Modeling Steps
  - Get Data
  - Initial Selections
  - Test Error/Link
  - Prelim Interrogation
  - Build Model
  - Validate Final Model
  - Combine Models
- Business Restrictions
- Summary

Review Model
Residual Analysis

- Re-check residuals to ensure appropriate shape



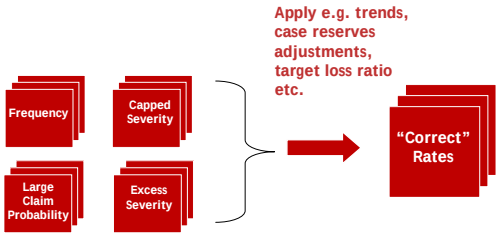


- Background
- Modeling Steps
  - Get Data
  - Initial Selections
  - Test Error/Link
  - Prelim Interrogation
  - Build Model
  - Validate Final Model
  - Combine Models
- Business Restrictions
- Summary

Combine Models for Required Coverage

- Objective is to determine a single rate and set of relativities for each coverage

Generalized Linear Models

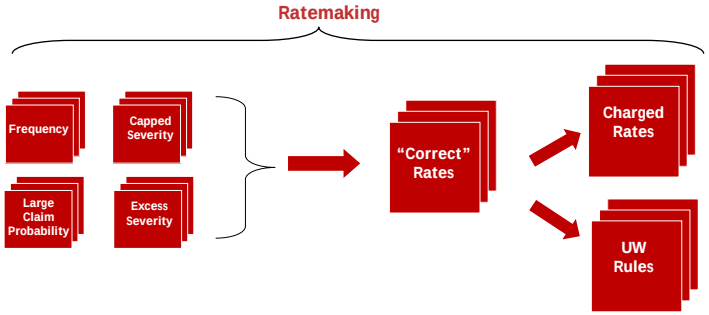


- Could alternatively model raw pure premiums instead, but frequency signals will be masked by severity noise

- Background
- Modeling Steps
  - Get Data
  - Initial Selections
  - Test Error/Link
  - Prelim Interrogation
  - Build Model
  - Validate Final Model
  - Combine Models
- Business Restrictions
- Summary

Ratemaking 101

- GOAL: determine the "best" price based on a variety of business considerations



- Once theoretical rates or pure premiums are known, analyst then incorporates changes related to other considerations
  - Structural differences (e.g., excluded factors, classification limitations, mixed rating algorithm)
  - Parameter differences (e.g., restricting relativities)
  - UW versus rating characteristics

- Background
- Modeling Steps
  - Get Data
  - Initial Selections
  - Test Error/Link
  - Prelim Interrogation
  - Build Model
  - Validate Final Model
  - Combine Models
- Business Restrictions
- Summary

Conclusion

- GLMs can be a powerful tool modeling tool with significant advantages over traditional techniques
- Basic steps same regardless of what is being modeled
  1. Get clean data
  2. Select an initial error structure, link function, model structure
  3. Test error structure/link function
  4. Preliminary investigation
  5. Build model structure
  6. Validate final predictive model
  7. Combine models
- Once "best" risk model built, can apply business judgment to determine proposed rates and underwriting rules

Thanks for coming, if you would like a copy of these slides:

- Give me your name/email after the session
- Call me at 210.826.2878
- Email me at geoff.werner@embamerica.com

GLM III will cover:

- Testing the link function
- The Tweedie distribution
- Splines-theory and practice
- Reference models
- Aliasing/near-aliasing
- Combining models across claim types
- Restrictions on models

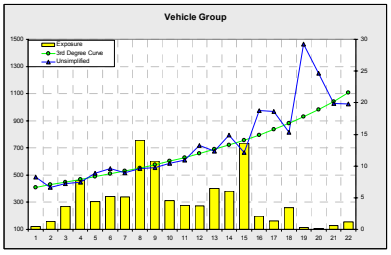
EMB America

Modeling-Fitting Curves (Variates)

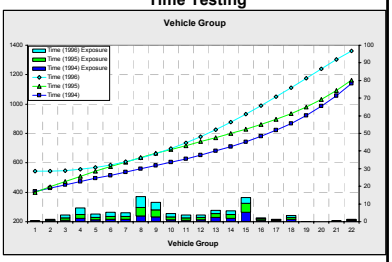
- Simplify trends in rating factors by fitting curves, splines, etc.

- Background
- Modeling Steps
  - Get Data
  - Initial Selections
  - Test Error/Link
  - Prelim Interrogation
  - Build Model
  - Validate Final Model
  - Combine Models
- Business Restrictions
- Summary

View Graphically



Time Testing



Parameters

Name	Value	Standard Error	Standard Error (%)	Weight
Vehicle Group Curve 1(OPoly(1))	0.2376	0.00676	2.8	18,779
Vehicle Group Curve 1(OPoly(2))	0.0112	0.00617	55.0	18,779
Vehicle Group Curve 1(OPoly(3))	0.0021	0.00521	253.9	18,779

Chi-Squared

Model	No Curve	Curve
Deviance	8,906.4460	9,020.2270
Degrees of Freedom	18,469	18,487
Scale Parameter	0.4822	0.4879
Chi Square Test		0.0%

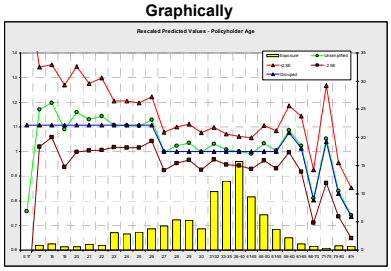
EMB America

Modeling-Grouping Rating Levels

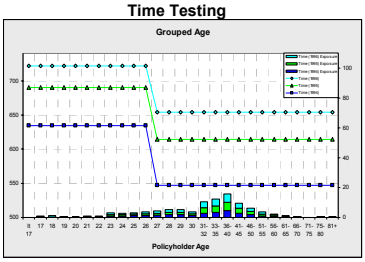
- Simplify trends in rating factors by grouping factor levels…

- Background
- Modeling Steps
  - Get Data
  - Initial Selections
  - Test Error/Link
  - Prelim Interrogation
  - Build Model
  - Validate Final Model
  - Combine Models
- Business Restrictions
- Summary

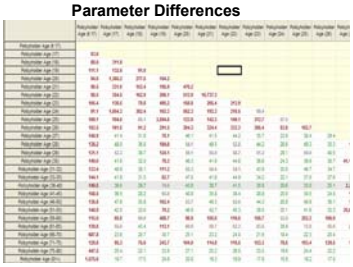
Graphically



Time Testing



Parameter Differences



Chi-Squared

Model	No Group	Grouped
Deviance	8,906.4420	8,935.1260
Degrees of Freedom	18,469	18,494
Scale Parameter	0.4822	0.4831
Chi Square Test		27.7%

EMB America