

Predictive Modeling for Workers Compensation

David Otto, FCAS, MAAA
EMB America LLC



Modeling Workers Compensation Risks

PURPOSE: To provide a technical discussion of solutions to the challenges associated with modeling workers compensation insurance.

OUTLINE

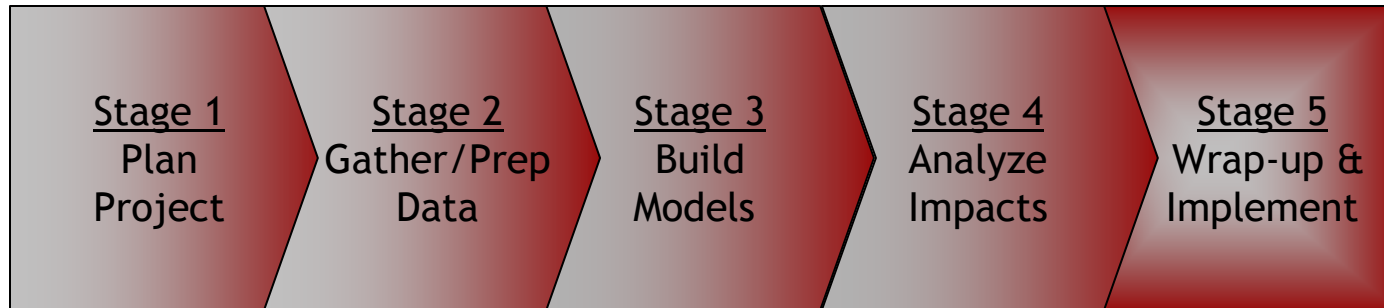
- ❖ Plan Project
- ❖ Gather Data
- ❖ Build Models
- ❖ Analyze
- ❖ Implement



EMB

Predictive Modeling Process

- Plan
- Gather Data
- Build Models
- Analyze
- Implement



1. Plan Project: establish scope, objectives, and requirements
2. Gather & Prepare Data: gather data and create necessary model files
3. Build Models: use historical data to build frequency and severity models including underlying development models for medical and indemnity (and possibly expense) losses and combine to form modeled pure premiums
4. Analyze Impacts: analyze the renewal, competitive, and profitability impacts of various proposals and finalize decisions
5. Wrap-Up: document decisions and communicate results



EMB

Plan Project

- Plan
- Gather Data
- Build Models
- Analyze
- Implement

GOAL: Establish scope, objective, and requirements

- ❖ Critical for proper project management
- ❖ Understand objectives and goals of all stakeholders
 - Timing
 - End product (most often an underwriting score but how is it expected to be used?)
 - Tiering
 - Schedule Rating - automated or recommended?
- ❖ Ensure all stakeholders understand benefits of predictive modeling over traditional techniques
- ❖ Important to involve underwriting personnel in the process because they will likely be the users of the final results



EMB

Gather & Prepare Data

- Plan
- Gather Data
- Build Models
- Analyze
- Implement

- ❖ Data preparation can consume over half of the time spent on a predictive modeling project
- ❖ Key data challenges with a workers compensation project
 - Volume
 - Underlying distribution
 - Dimensional dilution
 - Exposure base
 - Quality
 - Policy/loss matching
 - Null records
 - Dimensionality
 - Policy vs. claim rating variables
 - External data
 - Underwriting data



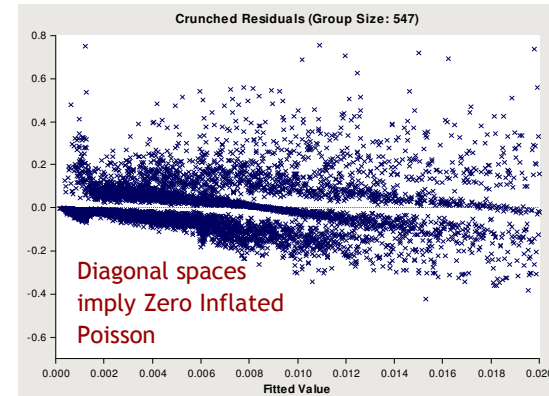
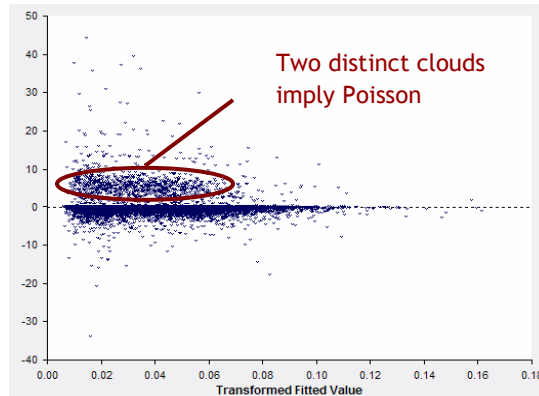
EMB

Data Challenge: Volume

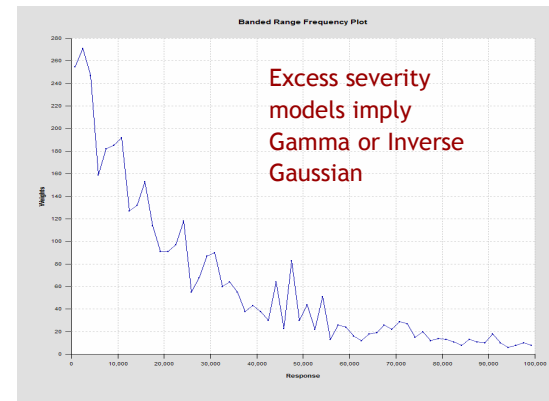
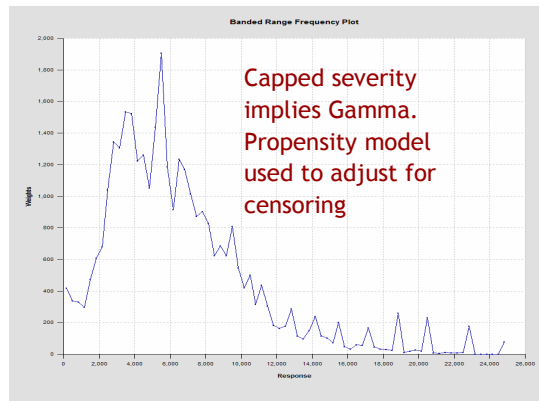
- Plan
- Gather Data
- Build Models
- Analyze
- Implement

Underlying Distribution

- Low frequency: fewer observations



- High severity: greater volatility

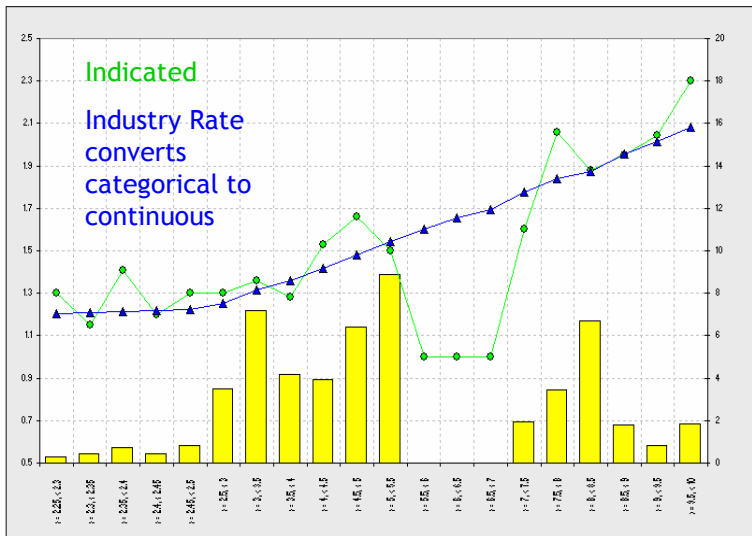


Data Challenge: Volume

- Plan
- Gather Data
- Build Models
- Analyze
- Implement

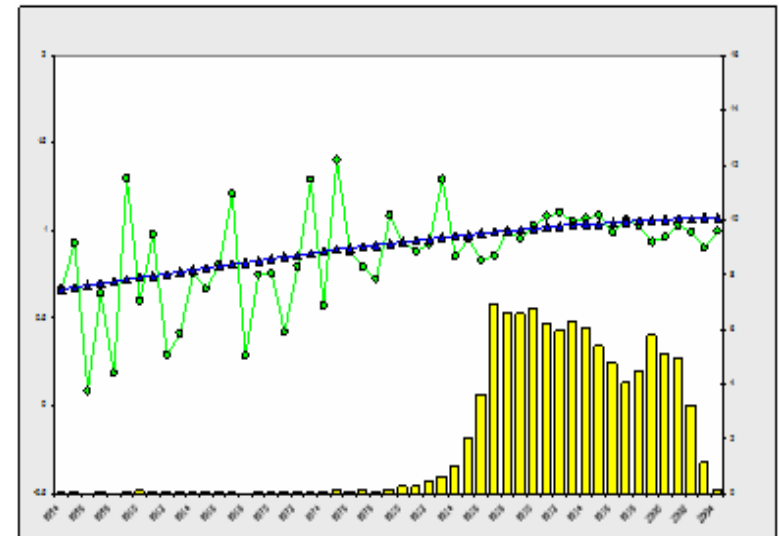
Dimensional Dilution

- Observations are spread thinly over multiple states
- Extrapolating beyond niche markets



Class Codes: Ranked by NCCI Rate

- Use indications from class codes with higher volumes
- Use NCCI rate for class codes with low volumes



Policy Year

- More years of data with a time element adds credibility without sacrificing trend

Data Challenge: Quality

- Plan
- Gather Data
- Build Models
- Analyze
- Implement

- ❖ Using new tools always seems to uncover previously undetected data problems that must be researched
- ❖ Typical issues
 - Bad data (e.g., 10 year old workers), especially for variables not used in rating
 - Poor linkage between losses and policy and class characteristics
 - No mapping of old groupings into new groupings (e.g., boundary changes)
 - Inconsistency between variables
 - Inconsistency within variables
 - Free flow versus set level data capture
- ❖ **Issues above are magnified with long-tailed lines** because of the need for a longer history of data

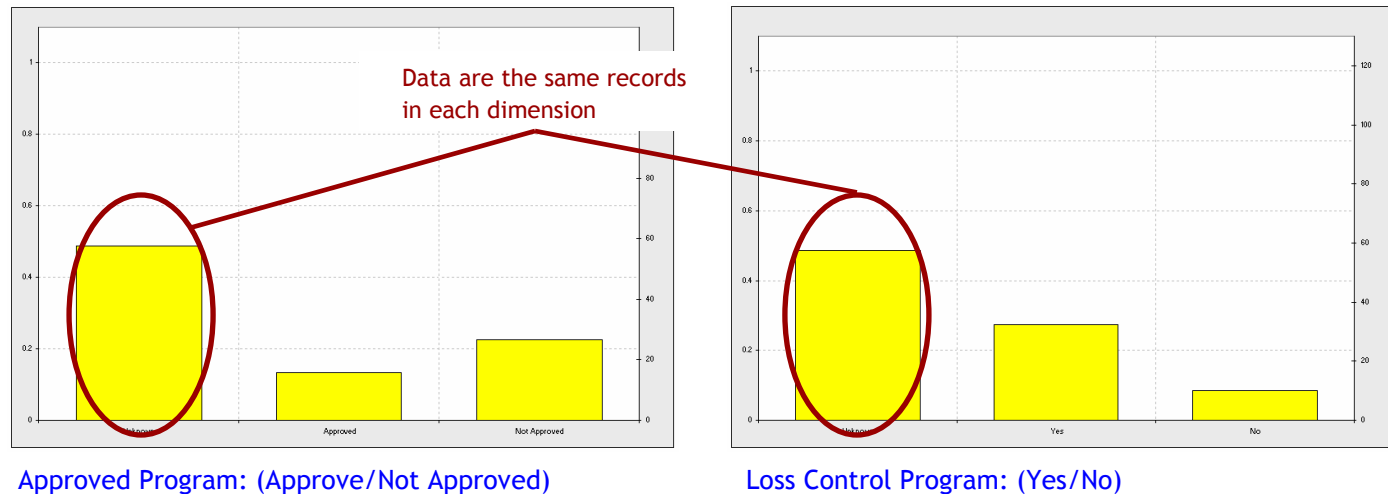


EMB

Data Challenge: Quality

- Plan
- Gather Data
- Build Models
- Analyze
- Implement

- ❖ Null records prevalent in workers compensation data sets
 - Creates complex aliases



- ❖ Solution
 - Create appropriate model structure to isolate the effect of the missing data
 - Assign a parameter for unknown in one rating factor and group unknown with the base in all other rating factors



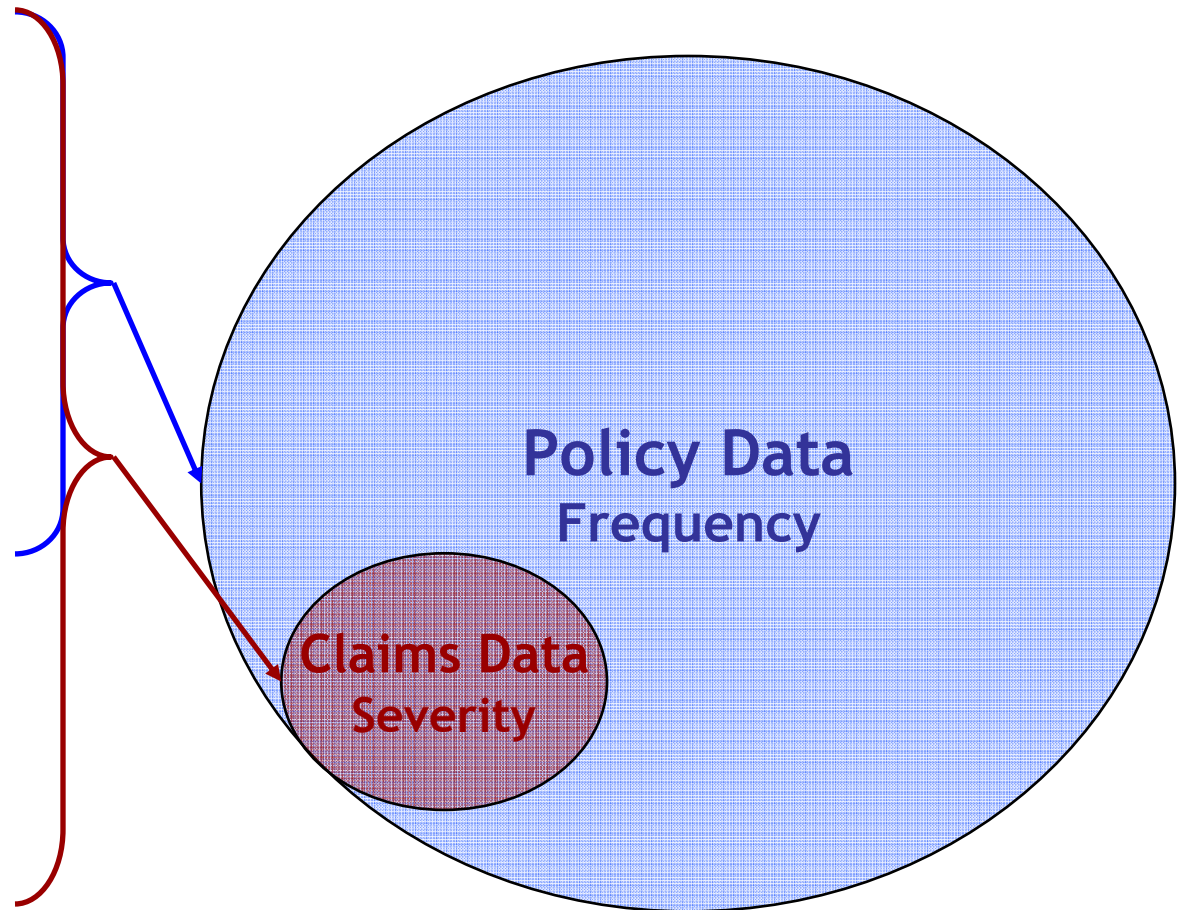
EMB

Data Challenge: Dimensionality

- Plan
- Gather Data
- Build Models
- Analyze
- Implement

- ❖ Workers compensation systems capture small amounts of policy information and a large amounts of claim information

Data:
Class code
Territory
Experience mod
Minimum premium
Employee count
Credit score
Agency
Years renewed
Company size
Premium discount
Schedule credit
Injury type
Injury description
Injury location
Age
Gender
Attorney
Report lag
Weekly wage
Marital status
...many more

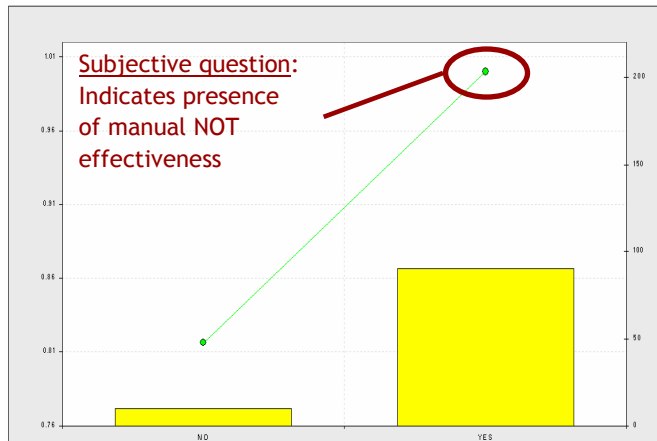


EMB

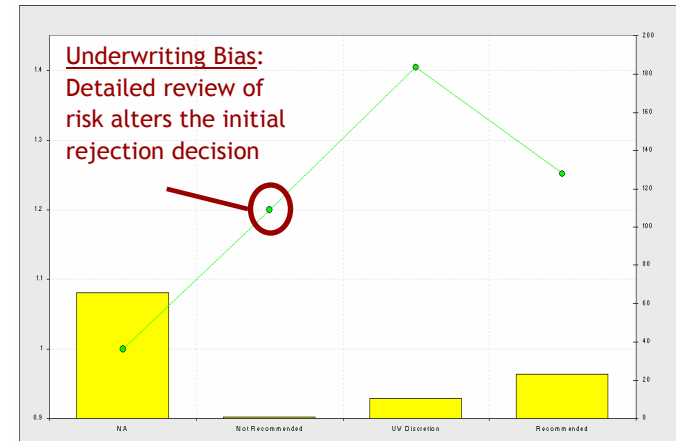
Data Challenge: Dimensionality

- Plan
- Gather Data
- Build Models
- Analyze
- Implement

- ❖ Policy data is often supplemented with external data and underwriting information
 - Incorporating external data
 - Cost
 - Time
 - Maintenance
 - Value
 - Incorporating underwriting / policy application data



Employee Manual



Underwriting Acceptance Criteria



EMB

Data Challenge: Dimensionality

❖ Credit Score

- Credit is a common external data source in personal lines
- Applicability to commercial lines?
- Problems:
 - Permanency:
 - Personal lines: credit sticks with you for life
 - Commercial lines: bad credit companies tend to go out of business and are often reborn under a new name
 - Linkage
 - Personal lines: personal info (i.e. ss number) provides match
 - Commercial lines: difficult to match
 - » Multiple mailing addresses
 - » Multiple company names within corporate structure
- Solutions:
 - Policy tenure may be a good proxy for credit
 - Try linking data with multiple variables to increase success rate
 - Phone number
 - Business name
 - FEIN

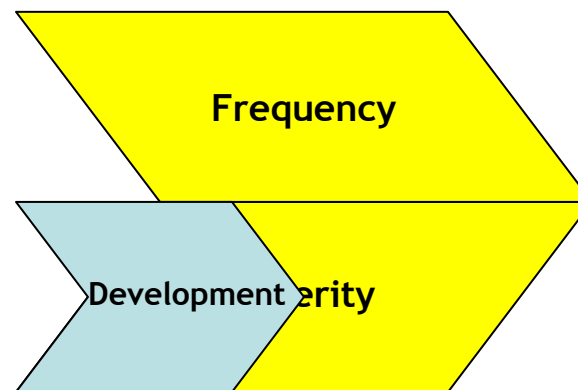


EMB

Build Models

- Plan
- Gather Data
- **Build Models**
- Analyze
- Implement

- ❖ Predictive modeling techniques are transportable to all types of insurance but commercial lines modeling has its own unique challenges that need to be properly addressed
- ❖ Key modeling challenges with a workers compensation project
 - Components
 - Development



EMB

Modeling Challenges: Components

- Plan
- Gather Data
- **Build Models**
- Analyze
- Implement

- ❖ A common misperception in commercial lines is to perform loss ratio modeling
- ❖ Modeling loss ratios is significantly inferior to pure premiums for both practical and theoretical reasons
 - Habit from traditional pricing methodology or necessity
 - Limitations of the predictive modeling software
 - Short cut



EMB

Modeling Challenges: Components

- Plan
- Gather Data
- **Build Models**
- Analyze
- Implement

❖ On-level Premium

- Historical data must be re-rated using current rates (i.e. extension of exposures)
 - Each risk must be re-rated (average current rate level factors will not work)
 - Workers compensation requires **REUNDERWRITING**
 - Credits/debits and special adjustments?
 - Underwriting judgments will not be consistent

❖ Extremely challenging task

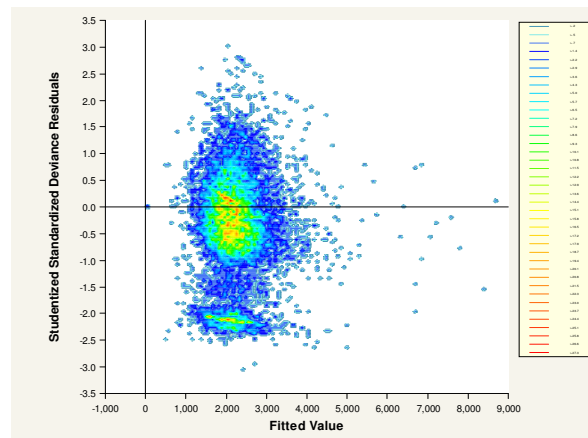


EMB

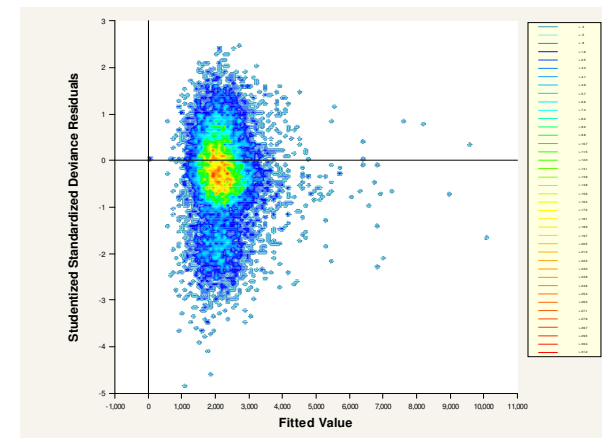
Modeling Challenges: Components

- Plan
- Gather Data
- Build Models
- Analyze
- Implement

- ❖ Loss Ratio Distributions
 - Tweedie
 - Gamma
- ❖ Heterogeneity within the response (frequency and severity components) require dispersion modeling techniques



Clusters of residuals due difference between frequency and severity signals



Dispersion modeling techniques improve fit

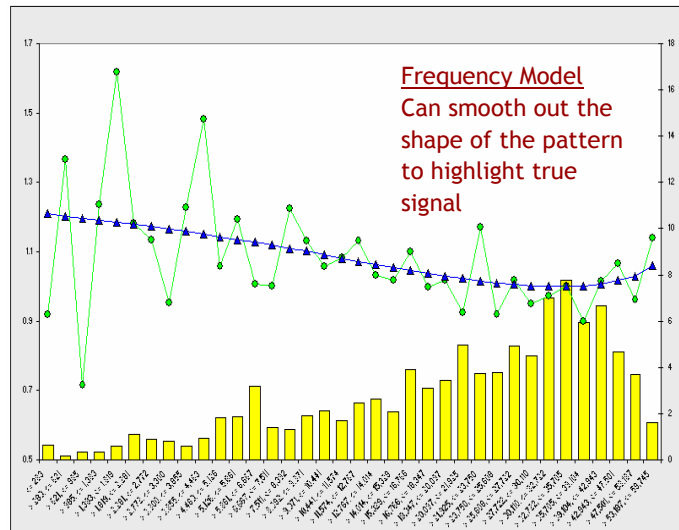
- ❖ No real intuitive sense of what the scale parameter response is in a dispersion model

Modeling Challenges: Components

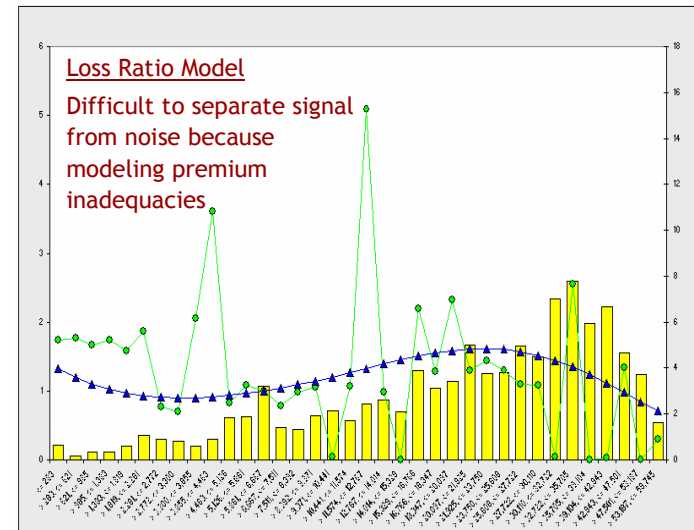
- Plan
- Gather Data
- Build Models
- Analyze
- Implement

Interpreting Loss Ratio Model Results

- No a priori expectation
- Shape of predictors will be dictated by how you are priced across all types of risk



County Population Densities



County Population Densities

- Problem is magnified as you begin to introduce new variables into the model

Modeling Challenges: Components

- Plan
- Gather Data
- Build Models
- Analyze
- Implement

- ❖ Implementing the results from Loss Ratio Model
 - Black box
 - Loss ratio modeling produces “black box” support
 - Explanatory power of GLM’s are greatly diminished
 - Buy-in from management, underwriting?
 - Short cut
 - Loss ratio model is not reusable
 - Entire dataset must be re-rated each time analysis is updated



EMB

Modeling Challenges: Development

- Plan
- Gather Data
- Build Models
- Analyze
- Implement

- ❖ Loss development within the predictive modeling process is not necessary if one of the following conditions is met:
 - Case incurred losses at the individual claim level are reasonably accurate
 - Common assumption in short-tailed lines
 - Each claim at a given lag is assumed to develop by the same proportion (i.e. the same development factor can be applied to each claim)
 - Common assumption in auto liability
 - Basic limits losses
 - Claims data matches policy data
- ❖ For personal lines the time variable is sufficient
- ❖ Workers compensation requires a more rigorous approach



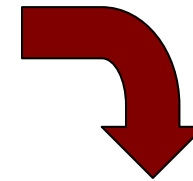
EMB

Modeling Challenges: Development

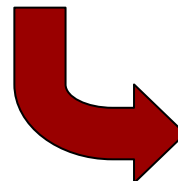
- Plan
- Gather Data
- Build Models
- Analyze
- Implement

- ❖ Traditional loss development: aggregate all claims in each cell within the historical triangle on a cumulative basis

<u>Claim</u>	<u>Accident Year</u>	<u>12</u>	<u>24</u>	<u>36</u>	<u>48</u>
000001	2002	0	1,000	1,000	5,000
000021	2002	50	50	50	50
000060	2002	0	0	0	250
000124	2003	300	500	500	
000328	2003	125	400	400	
000443	2003	0	100	2,000	
000776	2004	120	400		
000834	2004	100	200		
000942	2004	0	500		
001001	2005	0			
001075	2005	110			
001100	2005	100			



<u>Accident Year</u>	<u>12</u>	<u>24</u>	<u>36</u>	<u>48</u>
2002	50	1,100	2,150	7,450
2003	425	1,425	4,325	
2004	220	1,320		
2005	210			



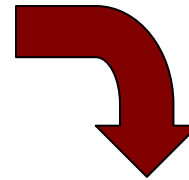
EMB

Modeling Challenges: Development

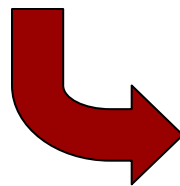
- Plan
- Gather Data
- Build Models
- Analyze
- Implement

- ❖ Traditional loss development goals:
 - Square the triangle
 - Forecast the tail

<u>Accident Year</u>	<u>12</u>	<u>24</u>	<u>36</u>	<u>48</u>
2002	50	1,100	2,150	7,450
2003	425	1,425	4,325	
2004	220	1,320		
2005	210			



<u>Accident Year</u>	<u>12</u>	<u>24</u>	<u>36</u>	<u>48</u>	<u>Ultimate</u>
2002	50	1,100	2,150	7,450	
2003	425	1,425	4,325	14,987	
2004	220	1,320	3,385	11,729	
2005	210	1,162	2,979	10,324	



- ❖ Estimating aggregate reserves does not produce a solution for allocating reserves to individual claims



EMB

Modeling Challenges: Development

- Plan
- Gather Data
- Build Models
- Analyze
- Implement

❖ GLM's can be used to develop individual claims to ultimate

<u>Claim</u>	<u>Accident Year</u>	<u>12</u>	<u>24</u>	<u>36</u>	<u>48</u>
000001	2002	0	1,000	1,000	5,000
000021	2002	50	50	50	50
000060	2002	0	0	0	250
000124	2003	300	500	500	
000328	2003	125	400	400	
000443	2003	0	100	2,000	
000776	2004	120	400		
000834	2004	100	200		
000942	2004	0	500		
001001	2005	0			
001075	2005	110			
001100	2005	100			



<u>Claim</u>	<u>Accident Year</u>	<u>Age</u>	<u>Amount</u>
000001	2002	12	0
000001	2002	24	1,000
000001	2002	36	1,000
000001	2002	48	5,000
000021	2002	12	50
000021	2002	24	50
000021	2002	36	50
000021	2002	48	50
000060	2002	12	0
000060	2002	24	0
000060	2002	36	0
000060	2002	48	250
000124	2003	12	300
000124	2003	24	500
000124	2003	36	500
000328	2003	12	125
000328	2003	24	400
000328	2003	36	400



EMB

Modeling Challenges: Development

- Plan
- Gather Data
- Build Models
- Analyze
- Implement

Development Challenges

- Functional relationship between variance and mean
 - Poisson
 - Binomial/Gamma process
 - Zero Inflated model
- Functional relationship between characteristics and incremental payment
 - Multiplicative
 - Additive
 - Logit
 - Non linear
- Squaring up the triangle on individual claims means that reserves for IBNR claims are not estimated
 - Workers compensation reporting lags have regulatory guidelines



EMB

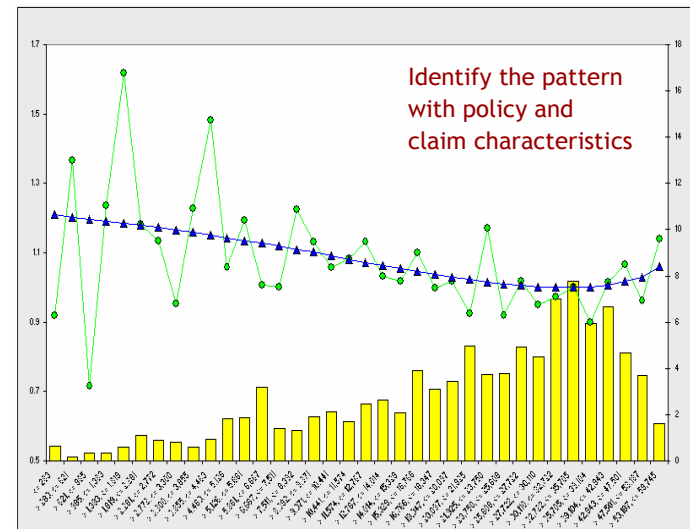
Modeling Challenges: Development

- Plan
- Gather Data
- Build Models
- Analyze
- Implement

Development Advantages

- Loss development enables the modeler to use claim characteristics as a means to separate signal from noise

WC Data:
Class code
Territory
Experience mod
Minimum premium
Employee count
Credit score
Agency
Years renewed
Company size
Premium discount
Schedule credit
Injury type
Injury description
Injury location
Age
Gender
Attorney
Report lag
Weekly wage
Marital status
...many more

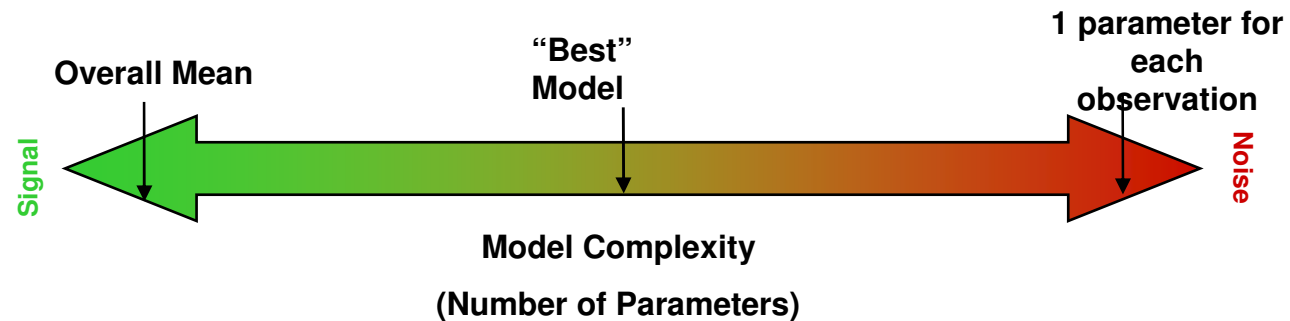


Without claim characteristics, the underlying signal will be difficult to identify

Modeling Challenges: Development

- Plan
- Gather Data
- Build Models
- Analyze
- Implement

- ❖ Once modeled, policy characteristics are then used to translate the smoothed patterns using information available when the policy is written

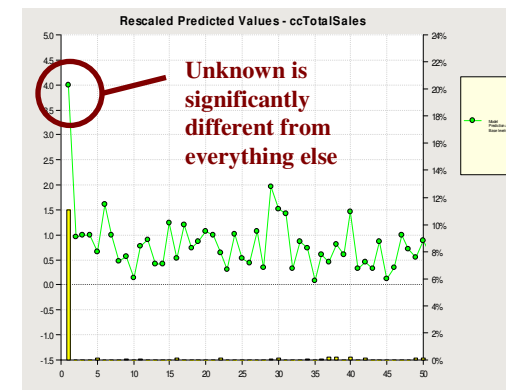
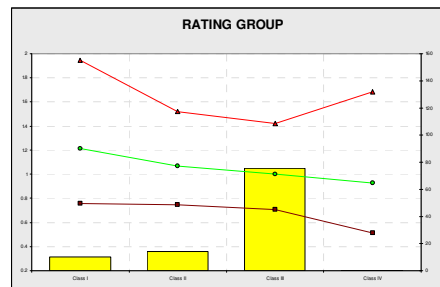


EMB

Analyze Impacts

Challenges:

- ❖ Traditional qualitative tests tend not to be as conclusive in commercial lines:
 - Standard error test leans towards underfitting the model
 - Chi-Square tests leans towards overfitting the model
 - Tripped up on Unknown levels



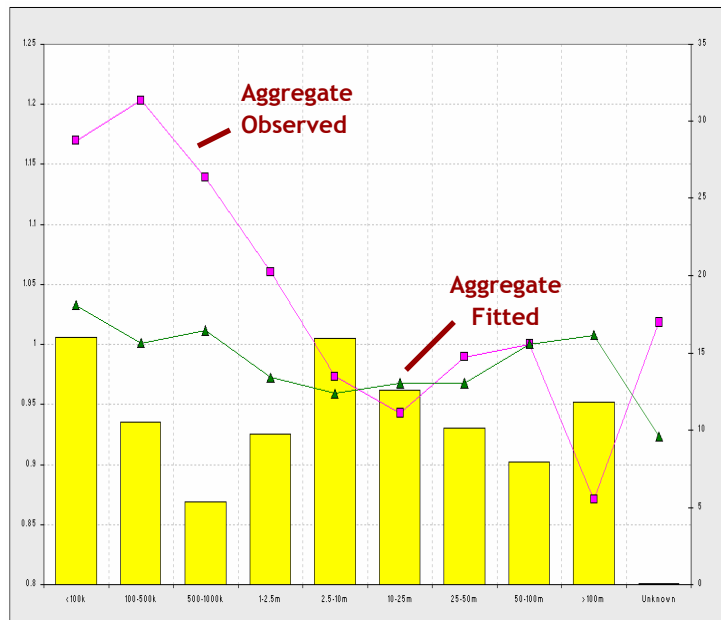
- ❖ Solution: Rely on additional diagnostics
 - Balance
 - Bias
 - Lift
 - Retrospective

Analyze Impacts

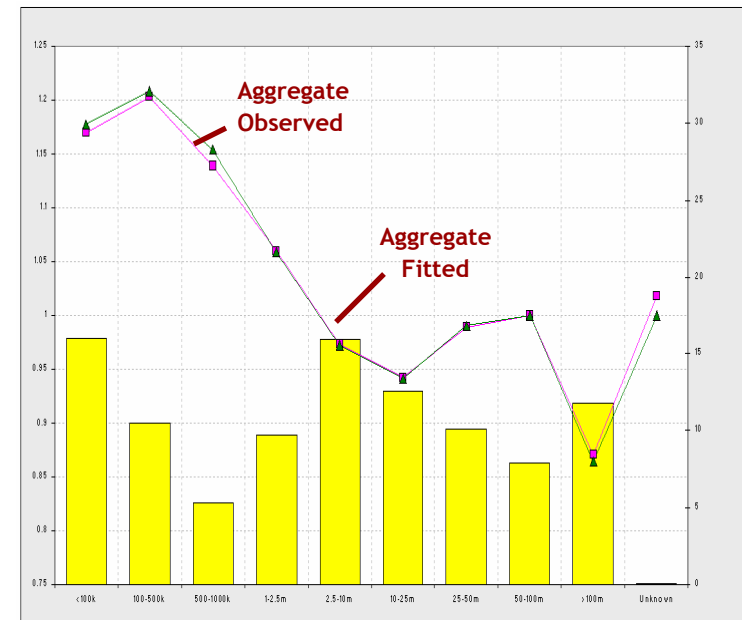
- Plan
- Gather Data
- Build Models
- Analyze
- Implement

Balance:

- Aggregate fitted results should be close to aggregate observed data across data dimensions



Account Size: Out of balance, too low for smaller accounts



Account Size: Used a more predictive model structure

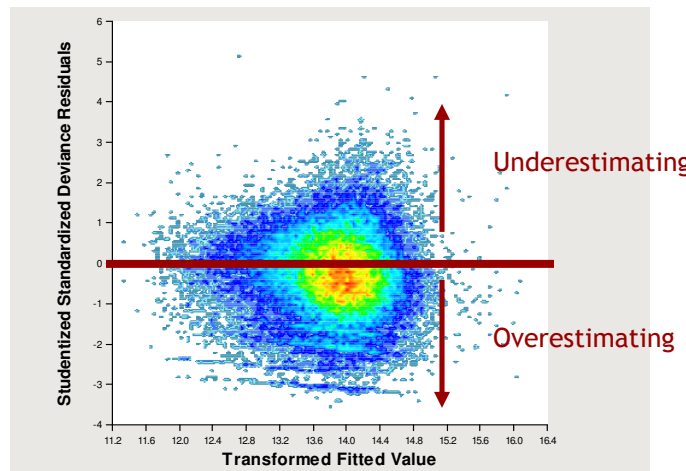
Analyze Impacts

- Plan
- Gather Data
- Build Models
- Analyze
- Implement

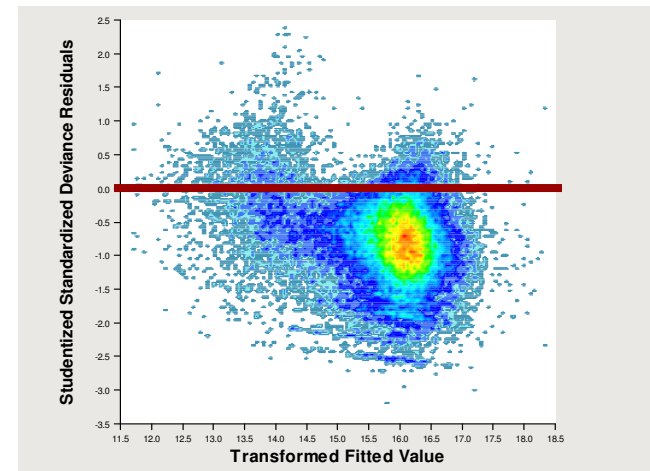
❖ Bias:

- Compare observed versus fitted data across all accounts

Example: No bias



Example: Overestimating



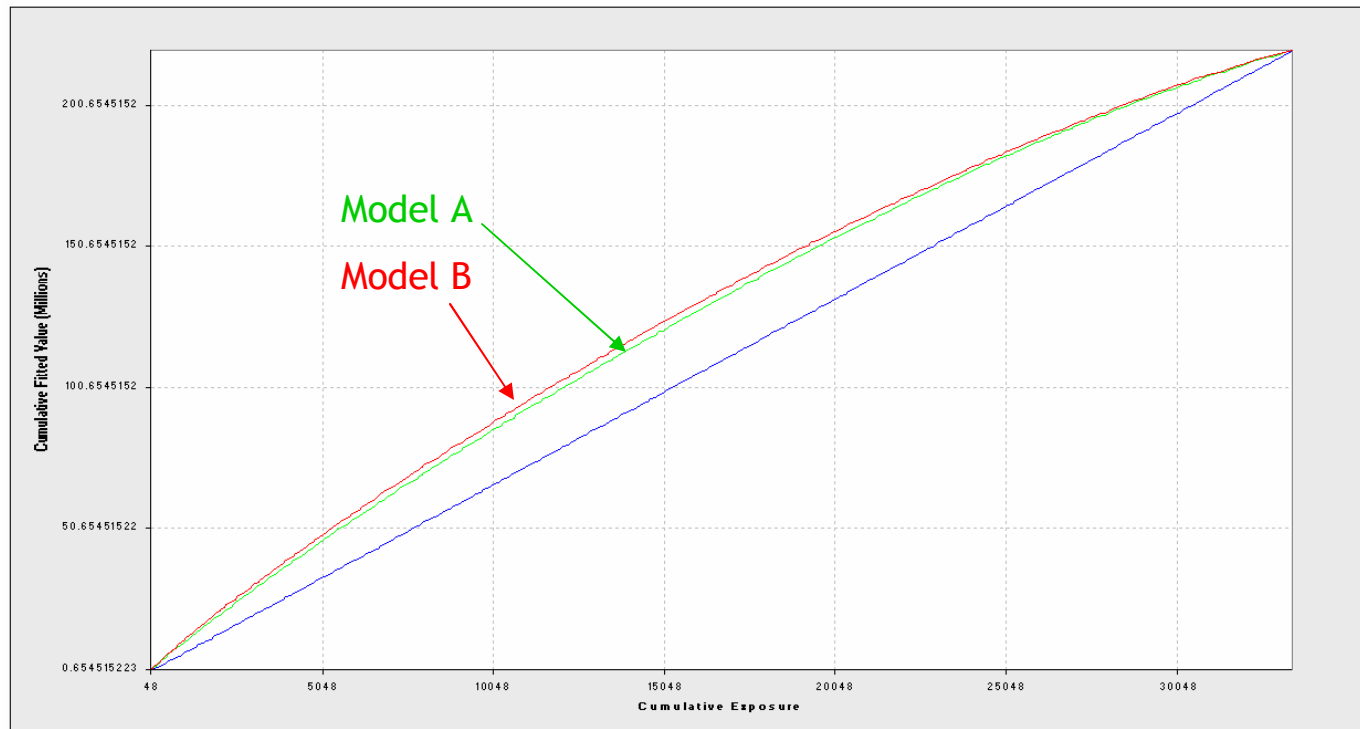
- Similar to balance but now focus is on individual accounts

Analyze Impacts

- Plan
- Gather Data
- Build Models
- Analyze
- Implement

▣ Lift

- Evaluate and quantify overall performance of selected model structures



- Model B is producing a greater lift than Model A



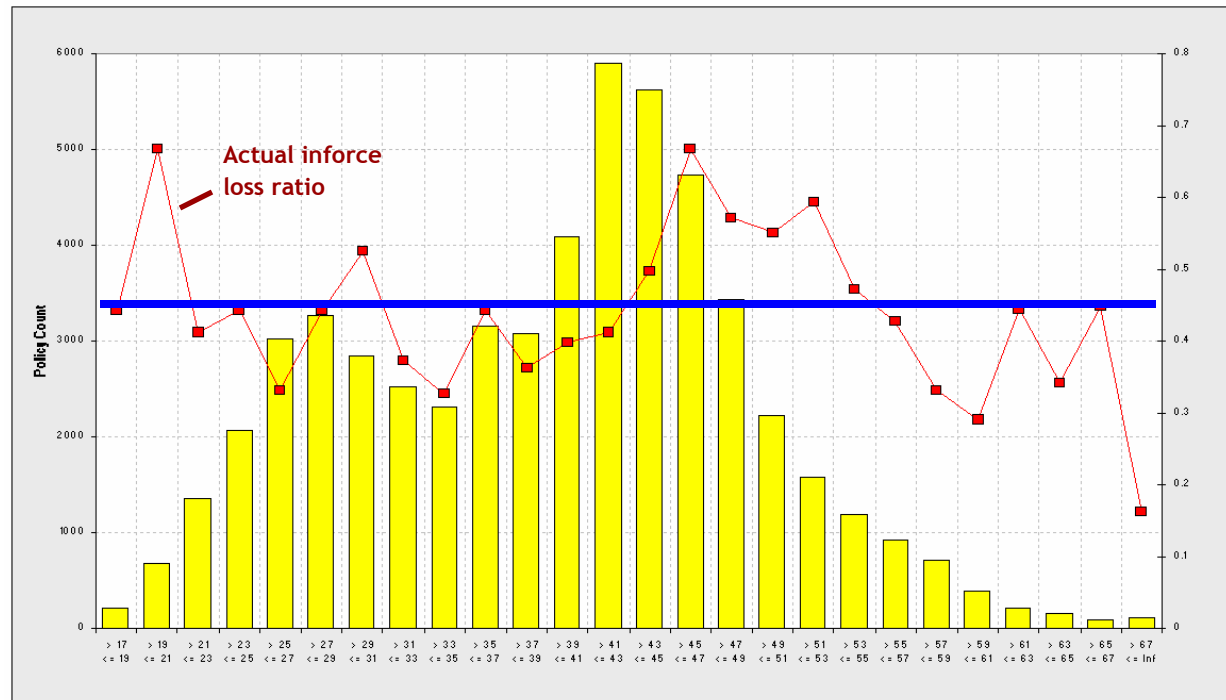
EMB

Analyze Impacts

- Plan
- Gather Data
- Build Models
- Analyze
- Implement

Retrospective:

- Test performance on new data that was not used to build model



Proposed Underwriting Score:

- Flat loss ratio from the latest inforce data set
- Volatility in loss ratio driven by immature data



EMB

Implement

- Plan
- Gather Data
- Build Models
- Analyze
- Implement

- ❖ Advanced techniques and technology enable the analyst to look at more explanatory variables than previously imagined
 - **Results:** Indications from predictive models will introduce factors not currently used in rating and relativities for existing factors that may be significantly different from current relativities
 - **Dilemma:** How to incorporate indications into existing rating plan?
 - Systems constraints
 - Tradition
 - Agency resistance
 - Market perception
 - Regulatory considerations



EMB

Implement

- Plan
- Gather Data
- Build Models
- Analyze
- Implement

- ❖ Consistency between underwriting and pricing
 - Underwriting emphasis is needed because of the uniqueness of each exposure
 - Overlap in qualitative assessments?
- ❖ Moving from yes/no underwriting scores to rate refinement
 - Underwriting scores help cherry-pick but do not establish the right rate for every risk
 - Either through rating or underwriting (example: tiers)
- ❖ Buy-in of distribution channels
 - Independent agents
 - MGA's
- ❖ Sizing up the competition
 - Industry focus on underwriting refinement makes it difficult to evaluate competitors

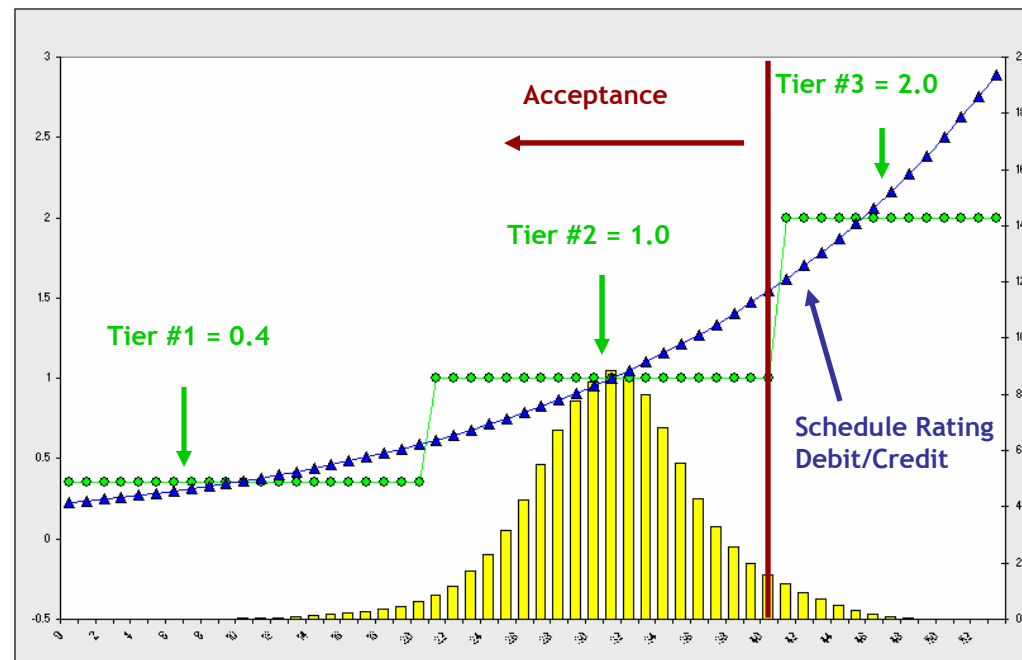


EMB

Implement

- Plan
- Gather Data
- Build Models
- Analyze
- **Implement**

- ❖ Scoring solutions allow multiple approaches toward implementation
 - Accept/Reject criteria
 - Tiering
 - Schedule Rating



Underwriting Score: Captures difference between indications and manual rate



EMB

Wrap-up

- Plan
- Gather Data
- Build Models
- Analyze
- Implement

GOAL: Document decisions and communicate results

- ❖ Create required filing support
- ❖ Decisions made during the review will need to be documented in accordance with actuarial standards
- ❖ Prioritize future actions

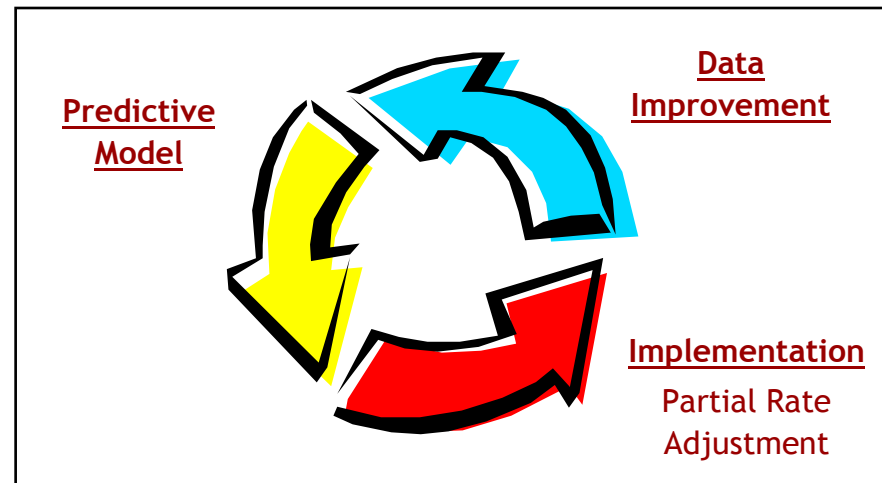


EMB

Wrap-up

- Plan
- Gather Data
- Build Models
- Analyze
- Implement

- ❖ Success in the personal lines marketplace has been widely publicized
- ❖ Important to remember that transition to current level of sophistication in the personal lines marketplace took many years
- ❖ Can get a lot of value through small incremental steps



EMB



EMB America LLC

Mission:

EMB America seeks to help our clients solve complex problems and identify opportunities by providing the appropriate blend of value-added consulting and state-of-the-art software. In so doing, EMB America strives to develop long-term relationships with clients and be the consulting firm of choice for the business community.

For information:

- ❖ Phone: 858.793.1425
- ❖ Email: inquiries@embamerica.com
- ❖ Website: www.embamerica.com

