

Frequency and Severity vs. Loss Cost Modeling

CAS 2012 Ratemaking and Product Management Seminar

March 2012

Philadelphia, PA

Alietia Caughron, PhD

Homesite Insurance Group



Antitrust Notice

- The Casualty Actuarial Society is committed to adhering strictly to the letter and spirit of the antitrust laws. Seminars conducted under the auspices of the CAS are designed solely to provide a forum for the expression of various points of view on topics described in the programs or agendas for such meetings.
- Under no circumstances shall CAS seminars be used as a means for competing companies or firms to reach any understanding – expressed or implied – that restricts competition or in any way impairs the ability of members to exercise independent business judgment regarding matters affecting competition.
- It is the responsibility of all seminar participants to be aware of antitrust regulations, to prevent any written or verbal discussions that appear to violate these laws, and to adhere in every respect to the CAS antitrust compliance policy.

Agenda

- Motivation
- Example

Motivation

- ❑ Breaking a problem into components
- ❑ Considering two separate questions
 1. Is there a claim?
 - Majority of policies have zero losses.
 - Frequency
 2. If there is a claim, how large is it?
 - Policies with non-zero losses are skewed.
 - Severity
- ❑ Versus, considering a compound distribution

Approach

- ❑ Data filtering, reconciliation, exploration
- ❑ Separate data into train & test 50/50
- ❑ Build model(s) on training data set, including main effects and any interactions
 - Significant effort goes into grouping levels using p-values, confidence intervals. Even at this stage, there is a balance between statistical results and rating, underwriting, or IT constraints.
 - Also consider AIC, BIC, lift curves. Balance with parsimony.

Approach

- ❑ Evaluate stability of selected variables, grouped levels, interactions using test data
- ❑ Evaluate model lift, stability of indications
- ❑ Use entire data set to determine final parameters, indicated relativities
 - Frequency & severity: Multiply together relativities produced by each model
 - Loss cost, or pure premium: Relativities produced automatically

Model selection

- ❑ Error structure belongs to the exponential family
- ❑ Variance = $\phi V(\mu)$
 where $V(\mu) = \mu^p$, $\phi > 0$ indicates dispersion

Error Structure	Mean	Variance	P
Poisson	μ	$\phi\mu$	1
Tweedie	μ	$\phi\mu^p$	$1 < p < 2$
Gamma	μ	$\phi\mu^2$	2

Model selection

➤ Two component models, vs one model

Model Component	Frequency	Loss Cost, or Pure Premium	Severity
Dependent variable	Claim count / Exposure	Loss / Exposure	Loss / Claim Count
Response	# claims	Total losses	Total losses
Weight	Exposures	Exposures	# claims
Link	Log	Log	Log
Error structure	Poisson	Tweedie, with p estimated	Gamma
Variance Function	μ^1	μ^p , where p belongs to (1,2)	μ^2

Results

- ❑ Variables selected for separate frequency and severity models will usually differ
- ❑ Not only will the variables selected differ, but also their relative 'importance'
- ❑ For pure premium models, the resulting set of variables reflects the ones selected in frequency and severity
 - Important to estimate p and not leave it fixed at a default value of say, 1.5

Selected variables

Variable	Frequency	Severity	Pure Premium
1	✓	✓	✓
2	✓		✓
3	✓	✓	✓
4	✓		✓
5	✓	✓	✓
6	✓		✓
7		✓	✓
...			✓
15		✓	✓

Selected variables (sorted)

Order	Frequency	Severity	Pure Premium
1	1	14	
2	2	5	
3	3	7	
4	4	9	
5	5	12	
6	6	15	
7	7	8	
...			
14	na		
15	na		

Example parameter estimates

Levels	One-way			GLM				
	Freq	Sev	Freq x Sev = Pure Prem	Freq	Sev	Freq x sev	Pure Prem p=1.67	Pure Prem p=1.5
intercept	na	na	na	-3.64	8.63	5.00	5.39	5.38
Base[1]	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A	0.29	0.17	0.51[2]	0.20	0.11	0.31	0.31	0.31
B	0.73	0.50	1.60	0.32	0.18	0.50	0.49	0.48
C	0.45	0.30	0.89	0.34	0.21	0.55	0.57	0.58

[1]Results shown for only one variable.

[2] $0.51=(0.29+1)*(0.17+1)-1$

Example relativities

Levels	One-way[1]			GLM				
	Freq	Sev	Freq x Sev = Pure Prem	Freq	Sev	Freq x sev	Pure Prem p=1.67	Pure Prem p=1.5
intercept	na	na	na	2.64%	5,614	148	219	218
Base[1]	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
A	1.34	1.18	1.66 [2]	1.22	1.11	1.36	1.36	1.36
B	2.08	1.65	4.94	1.37	1.20	1.64	1.63	1.61
C	1.57	1.36	2.44	1.41	1.23	1.74	1.77	1.78

[1]Results shown for only one variable.

[2]1.66=exp(0.51)

Frequency and severity models

- Greater understanding of business
- Easier to communicate
- Option to include a variable in either frequency or severity

- Modeled pure premiums can be produced to facilitate offsets, will require more work

Pure premium model

- Requires only one model to build and maintain
- Automatically adjusts for 'cancellation' effects
- Simpler method to implement offsets

- Pure premium approach allows only a binary choice for the inclusion of a variable

Recommendation

- First time through, build frequency and severity models
- Assuming this is a model that requires regular updates:
 - First or second time through, build all three models and compare results: frequency, severity and pure premium
 - Going forward, you can then focus on pure premium until there has been a significant shift in your data

Goal

- ❑ Important to remember the overall goal: a ‘reasonable’ model that pulls information out of the historical experience in such a way that it is likely to be predictive of the future.

