Predictive Modeling Loss Assumptions:

What's the Impact?

CAS RPM Seminar

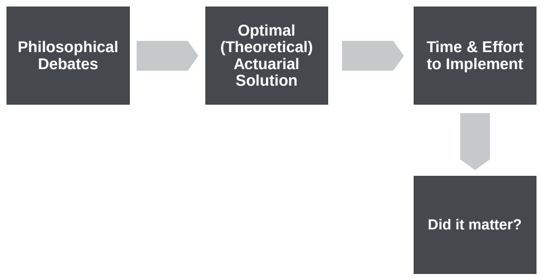
Eric Krafcheck, FCAS, MAAA, CSPA
Consulting Actuary
Katie Pipkorn, ACAS, MAAA
Associate Actuary
March 20, 2018

Overview

Milliman 2

Background

Impetus for Research



Milliman 4

Why care?

Most Theoretically Sound Methodology = Optimal Solution?
- Little guidance in actuarial literature
- Actuaries vs Data Scientists

Issue commonly ignored for sensitivity testing of model
- Scientific integrity

ASOPs
- ASOP 12, Risk Classification
  - ✓ No guidance
- ASOP 43, Unpaid Claim Estimates, §3.6.1
  - ✓ *"The actuary should consider methods or models for estimating unpaid claims that, in the actuary's professional judgment, are appropriate…The actuary should consider whether a particular method or model is appropriate in light of the purpose, constraints, and scope of the assignment."*
  - ✓ Scope: "…exclusive of estimates developed solely for ratemaking purposes."

Milliman 5

Why care? (cont.)

ASOPs
- ASOP 53, Estimating Future Costs for Prospective Property/Casualty Risk Transfer and Risk Retention (effective 8/1/18)
  - ✓ *§3.5: "The actuary should use methods or models, along with reasonable assumptions, that, in the actuary's professional judgment, have no known significant bias in the aggregate relative to the intended measure."*
  - ✓ *§3.8.2: "The actuary should consider adjusting historical data using methods or models, along with reasonable assumptions, that, in the actuary's professional judgment, reflect the ultimate value of the loss and loss adjustment expense. The actuary should also consider the following:*
    - *a) The coverage being evaluated;*
    - ***b) The type of analysis (such as overall future cost level analysis or risk classification analysis); and***
    - *c) The differences between the future period and the historical conditions under which the historical claims occurred, the claims were adjusted, and the claims reserves were set."*
  - ✓ *§3.8.3: "The actuary should consider past and prospective changes in claim costs, claim frequencies, exposures, and premiums."*

Milliman 6

Aggregate vs Individual Risk-Level Analyses

Traditionally, loss development is an aggregate analysis:

$$e.g.\ \underset{\substack{Incurred\ Loss \\ AY\ 2017}}{100{,}000}\ \ x\ \ \underset{12-Ult\ CDF}{2.742}\ \ =\ \ \underset{\substack{Ultimate\ Loss \\ AY\ 2017}}{274{,}200}$$

Predictive modeling analyses measure the relational and relative differences between risks based on risk characteristics

- Losses developed so relationships between risk characteristics are not distorted
- Total unpaid claim liability for LOB is irrelevant

Milliman 7

Loss Development Triangle

Accident Year	12-24	24-36	36-48	48-60	60-72	72-84	84-96	96-108	108-120
	Months of Development								
2008	1.667	1.405	1.200	1.087	1.019	1.006	1.000	1.000	1.000
2009	1.719	1.527	1.149	1.078	1.023	1.000	1.000	1.000	
2010	1.371	1.696	1.183	1.035	1.015	1.000	1.002		
2011	1.700	1.471	0.974	1.023	1.004	1.005			
2012	1.689	1.260	1.222	1.035	1.021				
2013	1.304	1.557	1.060	1.081					
2014	2.088	1.235	0.959						
2015	1.905	1.576							
2016	1.417								
Avg	1.651	1.466	1.107	1.056	1.017	1.003	1.001	1.000	1.000
Wtd Avg	1.641	1.425	1.089	1.055	1.017	1.003	1.001	1.000	1.000
Wtd Avg L3	1.768	1.386	1.061	1.046	1.014	1.002	1.001		
Selected	1.641	1.425	1.089	1.055	1.017	1.003	1.001	1.000	1.000
CDF	2.742	1.670	1.172	1.076	1.020	1.003	1.001	1.000	1.000

Milliman 8

Loss Development Triangle

Includes provision for:

1. Development on open claims
2. Development on closed claims (reopened claims)
3. Incurred but not reported claims
4. Reported but not recorded claims

IBNER (1–2)

"Pure" IBNR (3–4)

Selected	1.641	1.425	1.089	1.055		1.003	1.001	1.000	1.000
CDF	2.742	1.670	1.172	1.076	1.020	1.003	1.001	1.000	1.000

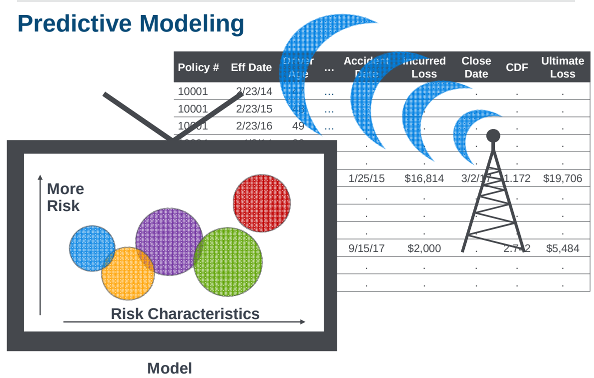
Milliman 9

Modeling Dataset

Policy #	Eff Date	Driver Age	…	Accident Date	Incurred Loss	Close Date	CDF	Ultimate Loss
10001	2/23/14	47	…	.	.	.	.	.
10001	2/23/15	48	…	.	.	.	.	.
10001	2/23/16	49	…	.	.	.	.	.
10004	4/2/14	30	…	.	.	.	.	.
10004	4/2/15	31	…	.	.	.	.	.
10005	11/28/14	62	…	1/25/15	$16,814	3/2/17	1.172	$19,706
10005	11/28/15	63	…	.	.	.	.	.
10005	11/28/16	64	…	.	.	.	.	.
10009	8/24/16	20	…	.	.	.	.	.
10010	7/16/17	25	…	9/15/17	$2,000	.	2.742	$5,484
10011	4/24/15	42	…	.	.	.	.	.
10012	9/1/16	23	…	.	.	.	.	.

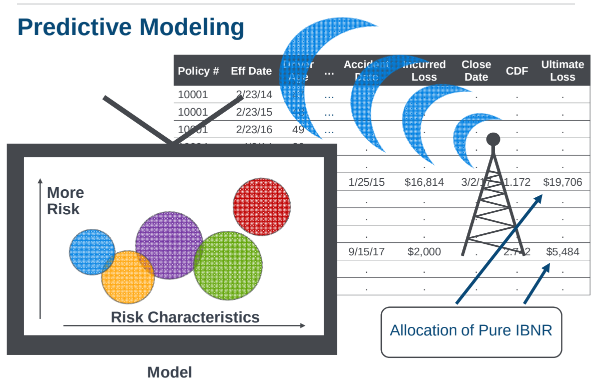
Milliman 10

Predictive Modeling



Milliman 11

Predictive Modeling



Milliman 12

Extreme Example

Assumptions

- **Company XYZ has 2 claim types: Type A and Type B**
  - ✓ Only non-youthful drivers have Claim Type A
  - ✓ Only youthful drivers have Claim Type B
- **Reporting**
  - ✓ Claim Type A: always reported within 12 months
  - ✓ Claim Type B: always reported between 12 and 24 months
  - ✓ For both, 50% of ultimate reported when claim is reported and remainder reported the following year
- **Severity**
  - ✓ Claim Type B's average severity is always 2x that of Claim Type A's
- **Frequency**
  - ✓ Both claim types occur with equal frequency

Extreme Example (cont.)

LDFs

- **12-24 MOD LDF: 4.00 (= [100% + 2 x (50%)] / 50%)**
- **24-36 MOD LDF: 1.50 (= [100% + 2 x (100%)] / [100% + 2 x (50%)])**

- **12-Ult CDF: 6.00 (= 4.00 x 1.50)**
- **24-Ult CDF: 1.50**

Extreme Example (cont.)

Claims Data

Claim #	AY	Claim Type	Incurred Loss	Open / Closed	CDF	Ultimate Loss
4	2017	A	1,000	Open	6.00	6,000
5	2017	A	1,000	Open	6.00	6,000
6	2017	A	1,000	Open	6.00	6,000
7	2016	B	2,000	Open	1.50	3,000
8	2016	B	2,000	Open	1.50	3,000
9	2016	B	2,000	Open	1.50	3,000
1	2016	A	2,000	Closed	1.50	3,000
2	2016	A	2,000	Closed	1.50	3,000
3	2016	A	2,000	Closed	1.50	3,000

Extreme Example (cont.)

Univariate Analysis:

- **Claim Type A Severity:**

$$\frac{(6,000+6,000+6,000+3,000+3,000+3,000)}{6}=4,500$$

- **Claim Type B Severity:**

$$\frac{(3,000+3,000+3,000)}{3}=3,000$$

- **Youthful Drivers Severity Relativity (Relative to Non-Youthful Drivers):**

$$\frac{3,000}{4,500}=0.667$$

Extreme Example (cont.)

Claims Dataset (with Corrected LDFs)

Claim #	AY	Claim Type	Incurred Loss	Open / Closed	Original CDF	Corrected CDF	Corrected Ultimate Loss
4	2017	A	1,000	Open	6.00	2.00	2,000
5	2017	A	1,000	Open	6.00	2.00	2,000
6	2017	A	1,000	Open	6.00	2.00	2,000
7	2016	B	2,000	Open	1.50	2.00	4,000
8	2016	B	2,000	Open	1.50	2.00	4,000
9	2016	B	2,000	Open	1.50	2.00	4,000
1	2016	A	2,000	Closed	1.50	1.00	2,000
2	2016	A	2,000	Closed	1.50	1.00	2,000
3	2016	A	2,000	Closed	1.50	1.00	2,000

Extreme Example (cont.)

Univariate Analysis (Corrected):

- **Claim Type A Severity:**

$$\frac{(2,000+2,000+2,000+2,000+2,000+2,000)}{6}=2,000$$

- **Claim Type B Severity:**

$$\frac{(4,000+4,000+4,000)}{3}=4,000$$

- **Youthful Drivers Severity Relativity (Relative to Non-Youthful Drivers):**

$$\frac{4,000}{2,000}=2.000$$

Adjustment Methodologies

Adjustment Methodologies

1	Control Variable Method
2	Unadjusted Loss Development Factor Method
3	Adjusted Loss Development Factor Method
4	Trend and Other Adjustments

Control Variable Method

Include time as an explanatory variable in model
- E.g. Policy year, accident year

Advantages
- Quick
- Easy to use
- No judgment required
- Accounts for both maturity and trend differences

Disadvantages
- Could possibly over-fit
- Doesn't allow judgment / expertise from user
- How to incorporate with machine learning algorithms?
- Limitations on validation design (e.g. last policy year in data)

Unadjusted Loss Development Factor Method

Use LDFs directly selected from loss development triangle

Advantages

- Easy to calculate
- Potentially readily available (ratemaking / reserving analyses)
- Doesn't require additional pure IBNR assumptions
- Allows user to incorporate judgment / expertise
- Can be used for machine learning techniques

Disadvantages

- Mismatch in allocation of IBNR
  - ✓ Pure IBNR allocated to reported claims
  - ✓ Closed and open claims may be over- or under-developed, respectively
- More time-consuming to implement than control variable method
- Does not account for trend differences

Milliman 22

Adjusted Loss Development Factor Method

Adjust LDFs to remove pure IBNR, effectively applying separate open and closed development factors to open and closed claims

Advantages

- Most actuarially sound method (theoretically)
- Properly allocates development
  - ✓ Pure IBNR excluded from analysis
  - ✓ Closed and open claims receive more appropriate development
- Allows user to incorporate judgment / expertise
- Can be used for machine learning techniques

Disadvantages

- Time-intensive
- May require multiple additional assumptions
  - ✓ Percent of development from newly reported claims
  - ✓ Allocation of development on closed and open claims
  - ✓ Are assumptions valid? How to verify?
- Does not account for trend differences

Milliman 23

Adjusted Loss Development Factor Method

Methodology

1	Select LDFs from loss triangle
2	Determine proportion of development related to pure IBNR
3	Remove pure IBNR from selected LDFs
4	Allocate remaining development from adjusted LDFs (from step 3) to open and closed claims
5	Calculate implied open and closed LDFs
6	Apply implied open LDFs to open claims and implied closed LDFs to closed claims

Milliman 24

Adjusted Loss Development Factor Method

Methodology

1. **Select LDFs from loss triangle**
2. Determine proportion of development related to pure IBNR
3. Remove pure IBNR from selected LDFs
4. Allocate remaining development from adjusted LDFs (from step 3) to open and closed claims
5. Calculate implied open and closed LDFs
6. Apply implied open LDFs to open claims and implied closed LDFs to closed claims

Milliman 25

Adjusted Loss Development Factor Method

Methodology

1. Select LDFs from loss triangle
2. **Determine proportion of development related to pure IBNR**
3. Remove pure IBNR from selected LDFs
4. Allocate remaining development from adjusted LDFs (from step 3) to open and closed claims
5. Calculate implied open and closed LDFs
6. Apply implied open LDFs to open claims and implied closed LDFs to closed claims

Milliman 26

Step 2: Determine proportion of development related to pure IBNR

Option 1: Directly measure

$$\text{\% of Dev. Attributable to Pure IBNR} = \frac{Losses\ on\ Newly\ Reported\ Claims_{MOD}}{Increm.\ Losses\ Reported\ Losses_{MOD-12\ to\ MOD}}$$

Option 2: Use reported claim count development pattern as proxy

- **Assumes loss development due to newly reported claims is proportional to claim count development**

Milliman 27

Adjusted Loss Development Factor Method

Methodology

1. Select LDFs from loss triangle
2. Determine proportion of development related to pure IBNR
3. **Remove pure IBNR from selected LDFs**
4. Allocate remaining development from adjusted LDFs (from step 3) to open and closed claims
5. Calculate implied open and closed LDFs
6. Apply implied open LDFs to open claims and implied closed LDFs to closed claims

Milliman 28

Step 3: Remove Pure IBNR from Selected LDFs

Subtract proportion of development attributable to pure IBNR from selected LDFs

		12-24	24-36	36-48	48-60
(1)	Selected LDF	3.500	2.100	1.250	1.115
(2)	Selected Reported Development Factor	2.250	1.050	1.010	1.002
(3)	% of Development Attributed to Pure IBNR	50.00%	4.55%	4.00%	1.74%
(4)	Adj LDF (Net of Pure IBNR)	2.250	2.050	1.240	1.113

(2) = From Claim Count Triangle Selections
(3) = [(2) – 1] / [(1) – 1]
(4) = [(1) – 1] * [1 – (3)] + 1

Milliman 29

Adjusted Loss Development Factor Method

Methodology

1. Select LDFs from loss triangle
2. Determine proportion of development related to pure IBNR
3. Remove pure IBNR from selected LDFs
4. **Allocate remaining development from adjusted LDFs (from step 3) to open and closed claims**
5. Calculate implied open and closed LDFs
6. Apply implied open LDFs to open claims and implied closed LDFs to closed claims

Milliman 30

Step 4: Allocate Remaining Development to Open / Closed Claims

1) Calculate development implied by LDFs from previous step

2) Select portion of development to allocate to open / closed claims

3) Multiply the implied development by the selected allocation

Step 4: Allocate Remaining Development to Open / Closed Claims

		12-24	24-36	36-48	48-60
(1)	Selected LDF	3.500	2.100	1.250	1.115
(2)	Selected Reported DF	2.250	1.050	1.010	1.002
(3)	% of Development Attributed to Pure IBNR	50.00%	4.55%	4.00%	1.74%
(4)	Adj LDF (Net of Pure IBNR)	2.250	2.050	1.240	1.113
(5)	Incurred Loss	1,000	3,500	7,400	9,250
(6)	Implied Development	1,250	3,675	1,776	1,045
(7)	% Allocated to Open	90%	95%	100%	100%
(8)	% Allocated to Closed	10%	5%	0%	0%
(9)	Development on Open	1,125	3,491	1,776	1,045
(10)	Development on Closed	125	184	0	0

(6) = [(4) – 1] x (5)   (9) = (6) x (7)

(7), (8) based on input from Claims and judgment   (10) = (6) x (8)

Adjusted Loss Development Factor Method

Methodology

1. Select LDFs from loss triangle
2. Determine proportion of development related to pure IBNR
3. Remove pure IBNR from selected LDFs
4. Allocate remaining development from adjusted LDFs (from step 3) to open and closed claims
5. **Calculate implied open and closed LDFs**
6. Apply implied open LDFs to open claims and implied closed LDFs to closed claims

Step 5: Calculate Implied Open and Closed LDFs

Implied Open LDF:

$$= \frac{Allocated\ Incr\ Open\ Dev_{MOD} + Incurred\ Loss\ on\ Open\ Claims_{MOD-12}}{Incurred\ Loss\ on\ Open\ Claims_{MOD-12}}$$

Implied Closed LDF:

$$= \frac{Allocated\ Incr\ Closed\ Dev_{MOD} + Incurred\ Loss\ on\ Closed\ Claims_{MOD-12}}{Incurred\ Loss\ on\ Closed\ Claims_{MOD-12}}$$

Select and smooth development patterns

- Verify implied development from Open and Closed LDFs reconciles to total implied development

Adjusted Loss Development Factor Method

Methodology

1	Select LDFs from loss triangle
2	Determine proportion of development related to pure IBNR
3	Remove pure IBNR from selected LDFs
4	Allocate remaining development from adjusted LDFs (from step 3) to open and closed claims
5	Calculate implied open and closed LDFs
6	**Apply implied open LDFs to open claims and implied closed LDFs to closed claims**

Trend Factors

Accounts for differences in cost-levels and / or claim frequencies

Advantages

- Easy to calculate
- Potentially readily available (ratemaking / reserving analyses)
- Allows user to incorporate judgment / expertise
- Can be used for machine learning techniques

Disadvantages

- More time-consuming to implement than control variable method
- By itself does not account for differences in maturity

Other Adjustment Techniques

Exposure / Weight Adjustments

- **For greener years, judgmentally adjust weight**
  - ✓ Reduces "credibility" of observations
  - ✓ Allows for incorporation of more recent experience
  - ✓ Does not correct for misallocation of IBNR

Allocation of IBNR to individual claim-level

- ✓ IBNR-to-case ratios, etc.

Other techniques?

Milliman 37

Research Results

Design

Evaluated 19 loss cost / severity models & 10 frequency models (GLMs)

- **Homeowners:**
  - ✓ Fire
  - ✓ Hail
  - ✓ Liability
  - ✓ Theft
  - ✓ Water
  - ✓ Wind / Lightning
- **Auto:**
  - ✓ Bodily Injury
  - ✓ Property Damage
  - ✓ Collision
  - ✓ Comprehensive

For each model, ran with various loss assumptions:

- **Unadjusted incurred loss and ALAE**
- **Unadjusted incurred loss and ALAE with Policy Year control variable**
- **Developed and trended loss and ALAE (with unadjusted LDFs)**
- **Developed and trended loss and ALAE (with pure-IBNR adjusted LDFs)**

Milliman 39

Design (cont.)

Compared results

- Diagnostics / Goodness-of-Fit measures
- Lift charts
- Distribution of change in predicted (relative to base scenario)
- Indicated estimates (i.e. relativities)

Areas assessed

- Variable selection process
- Fit
- Predictiveness
- Model estimates

Additional Details / Qualifications

Control variable

Maturity of datasets

Initial Hypotheses

1) Control variable method will over-fit

2) Most predictive model: Adj LDF Method

3) Differences most notable in longer-tailed coverages / perils / LOB

4) Differences most notable when less data available

Variable Selection

Variable Selection

Impact is minimal…unless you have thin data

- Considering Type III tests only
- Variable Selection Impact Distribution

	Loss Cost / Severity Models	Frequency Models
No Impact	42%	80%
Minimal Impact	42%	20%
Significant Impact	15%	0%

- ✓ Minimal impact = 1 or 2 variables potentially affected
- ✓ Significant impact = 3+ variables affected or convergence issues

- No pervasive trend by loss variable

Milliman

Fit

Deviance (% Change Relative to Base Scenario)

Incurred Loss

	All Models	Long-Tailed Models	Short-Tailed Models	Sufficient Data Models	Thin Data Models
Mean	3.9%	7.5%	1.9%	0.6%	8.4%
Standard Deviation	7.0%	9.9%	4.1%	4.0%	7.9%

*Excludes frequency models

Milliman 46

Deviance (% Change Relative to Base Scenario)

Policy Year Control Variable

	All Models	Long-Tailed Models	Short-Tailed Models	Sufficient Data Models	Thin Data Models
Mean	8.4%	20.4%	1.8 %	0.5%	19.0%
Standard Deviation	24.0%	39.4%	4.1%	4.0%	35.2%

*Excludes frequency models

Milliman 47

Deviance (% Change Relative to Base Scenario)

Trended + Developed (Unadjusted LDFs)

	All Models	Long-Tailed Models	Short-Tailed Models	Sufficient Data Models	Thin Data Models
Mean	2.5%	-4.5%	6.4%	4.5%	-0.1%
Standard Deviation	12.3%	18.6%	5.2%	5.4%	18.5%

*Excludes frequency models

Milliman 48

Other Goodness-of-Fit Measures

Scaled deviance, AIC / AICC / BIC, etc.

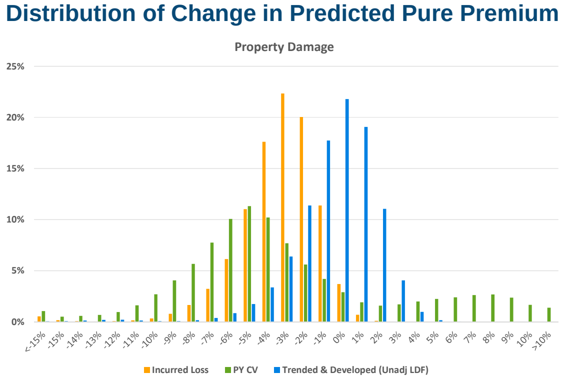
- Similar patterns / relationships

Type III

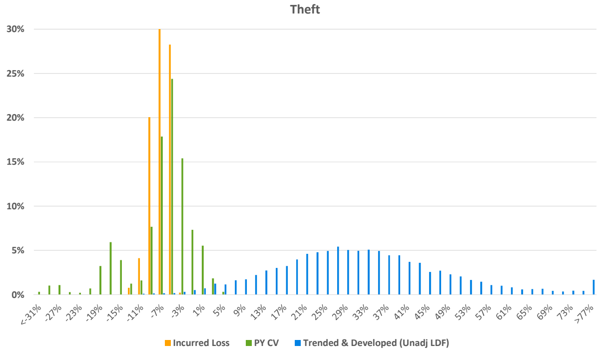
- No clear "winner"

Predictiveness

Distribution of Change in Predicted Pure Premium



Distribution of Change in Predicted Severity



Milliman 52

Gini Index

Compared relative to trended + developed (adjusted LDFs) model

- **Training and holdout bases**

No general consensus…

- **PY Control Variable**
  - ✓ Measured on a training basis: PY Control Variable performed better
  - ✓ Measured on a holdout basis: mixed
- **Trended + Developed (Unadjusted LDFs)**
  - ✓ Mixed results

…However

- **Relative measure tended to decrease from training to holdout bases for long-tailed / thin data models**

Milliman 53

Holdout Lift Charts

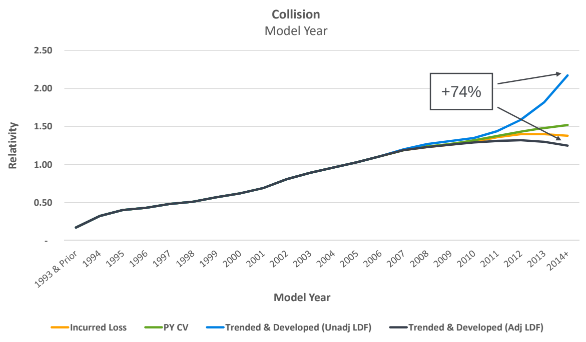
Compared on an SSE and "visual" basis

No general consensus

Milliman 54

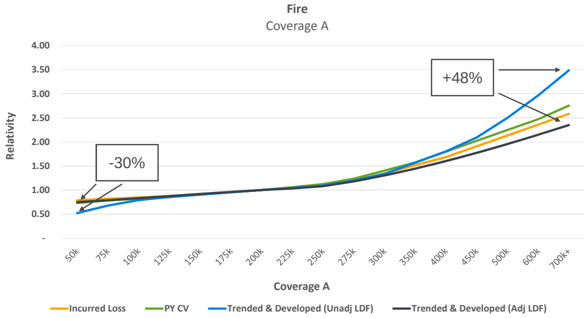
Model Estimates

Indicated Relativities



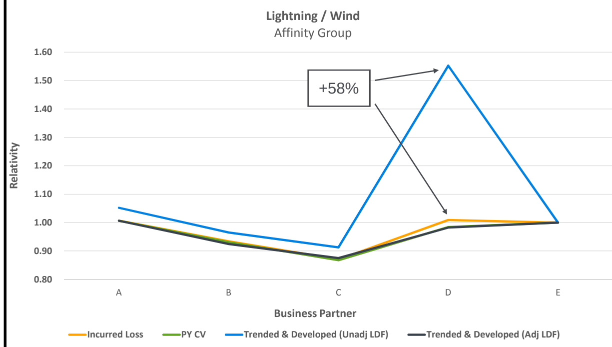
Milliman

Indicated Relativities



Milliman

Indicated Relativities



Milliman 58

Conclusions

59

Summary

1) Impact on variable selection is potentially minimal.

2) Impact varies by length of tail.

3) Impact varies by volume of data.

4) Potentially significant differences in predicted values and / or model estimates.

5) No clear "winner," but unadjusted LDFs tend to lead to more extreme results.

Conclusion: sensitivity testing important!

Milliman 60

Additional Considerations

Small sample size

Results impacted by reserving practices of Claims department

Data availability for LDFs

Potential for abuse

- Loss assumptions should not be selected to achieve a desired outcome (e.g. steeper credit curve, etc.)

Milliman

Thank you!

eric.krafcheck@milliman.com
(262) 796-3334

katie.pipkorn@milliman.com
(262) 923-3661