

A Quantum Leap in Benchmarking Unpaid Claims



Benchmarking Unpaid Claim Estimates

▪ **Benchmark:** A standard, or a set of standards, used as a point of reference for evaluating performance or level of quality. Benchmarks may be drawn from a firm's own experience, from the experience of other firms in the industry, or from legal requirements such as environmental regulations.

Source: businessdictionary.com

Milliman

2

Benchmarking Unpaid Claim Estimates

- Have you ever calculated an estimate of unpaid claims?
 - P&C (General) Insurance, any LOB or segment
 - For any reason, reserves, pricing, ERM, etc.
- Have you ever used a benchmark to help with your estimated unpaid claims or range of estimates?

Milliman

3

Benchmarking Unpaid Claim Estimates

Outline

- 1 Background
- 2 Analysis Summary
- 3 Model Limitations
- 4 Model Projections – Are they Unbiased?
- 5 Proposed Adjustments
- 6 Conclusions
- 7 Claim Variability Benchmarks

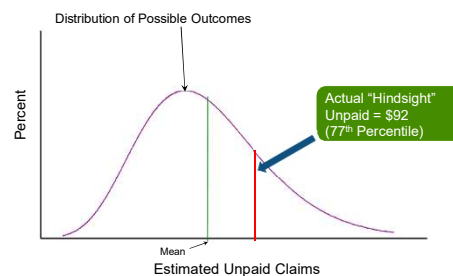
Milliman

4

Background

Hindsight Analysis

Hypothetical Unpaid Claim Distribution

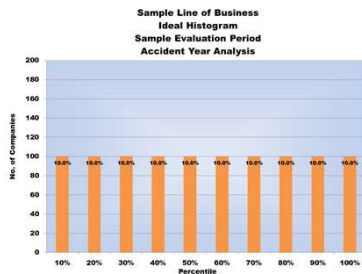


Milliman

6

A Quantum Leap in Benchmarking Unpaid Claims

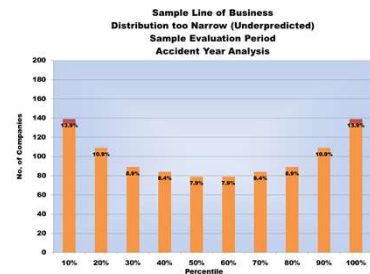
If Model is Correct...



Milliman

7

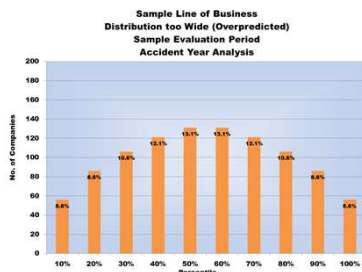
If Model Underestimates Distribution...



Milliman

8

If Model Overestimates Distribution...



Milliman

9

Background

Prior Research

Meyers & Shi

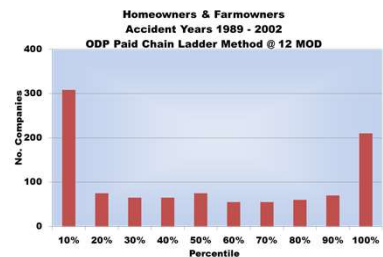
“...study suggests that there might be environmental changes that no single model can identify.”

“If this continues to hold, the actuarial profession cannot rely solely on stochastic loss reserve models to manage its reserve risk.”

Milliman

11

Leong, Wang & Chen



Leong, Jessica (Weng Kah), Shaun Wang, and Han Chen, “Back-Testing the ODP Bootstrap of the Paid Chain-Ladder Model with Actual Historical Claims Data,” CAS E-Forum, Summer 2012, 1-34.

Milliman

12

A Quantum Leap in Benchmarking Unpaid Claims

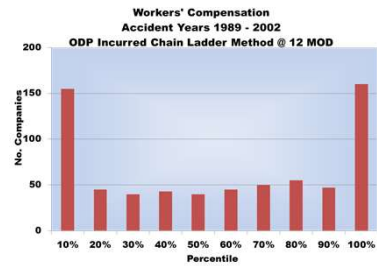
Leong, Wang & Chen

“...the popular ODP bootstrap of the paid chain-ladder method is underestimating reserve risk.”

“...the bootstrap model does not consider systemic risk, or, to put it another way, the risk that future trends in the claims environment – such as inflation, trends in tort reform, legislative changes, etc. – may deviate from what we saw in the past.”

Leong, Jessica (Weng Kah), Shaun Wang, and Han Chen, “Back-Testing the ODP Bootstrap of the Paid Chain-Ladder Model with Actual Historical Claims Data,” CAS E-Forum, Summer 2012, 1-34.

Leong, Wang & Chen



Leong, Jessica (Weng Kah), Shaun Wang, and Han Chen, “Back-Testing the ODP Bootstrap of the Paid Chain-Ladder Model with Actual Historical Claims Data,” CAS E-Forum, Summer 2012, 1-34.

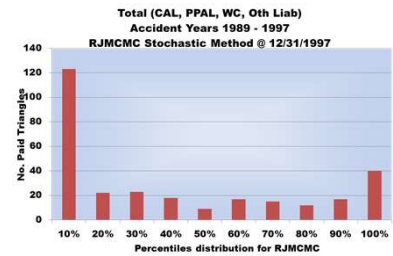
Leong, Wang & Chen

“...it appears that the incurred bootstrap model is also underestimating the risk of falling in these extreme percentiles.”

Note: This is not the same incurred ODP bootstrap model as described in the Shapland Monograph.

Leong, Jessica (Weng Kah), Shaun Wang, and Han Chen, “Back-Testing the ODP Bootstrap of the Paid Chain-Ladder Model with Actual Historical Claims Data,” CAS E-Forum, Summer 2012, 1-34.

Gremillet & Miehé



Gremillet, Marion, and Pierre Miehé, “Back-Testing the Reversible Jump Markov Chain Monte Carlo & further extensions,” ICA 1-38 (2013).

Gremillet & Miehé

“...it is core to have adjustments by actuaries prior to running the stochastic methods ‘automatically.’ ”

“Actuary in the box” dream for stochastic reserves valuation not yet happening

Gremillet, Marion, and Pierre Miehé, “Back-Testing the Reversible Jump Markov Chain Monte Carlo & further extensions,” ICA 1-38 (2013).

Analysis Summary

A Quantum Leap in Benchmarking Unpaid Claims

Comparison of Analyses

Item	Meyers & Shi	Leong, Wang & Chen	Gremillet & Miehe	Shapland
Data	50 Companies	21 (MPL) to 78 (PPAL) Companies	?	1,679 Companies
Evaluations	1	11	5	9
Models	2	2	3	8
Lines of Business	1	9	4	16
Triangle Sets	50	~4,850	296	30,707

Milliman

31

Analysis Details

- **ODP Bootstrap**
 - Paid Chain Ladder
 - Incurred Chain Ladder
 - Paid Bornhuetter-Ferguson
 - Incurred Bornhuetter-Ferguson
 - Paid Cape Cod
 - Incurred Cape Cod
 - Weighted
- **Mack Bootstrap**
 - Paid Chain Ladder

Milliman

32

Analysis Details

- **Beginning Data**
 - NAIC Schedule P – 4,796 Companies (& Groups)
 - Remove all triangles without 10 years of data (Paid, Incurred, etc.)
 - Other data quality tests → “quality data”
 - Test whether next 9 years are identical → “complete data”
- **Test Data**
 - Total of 75,000+ LOBs with “quality data”
 - 1,679 Companies with at least 1 Schedule P LOB of “complete data”
 - Total of 30,707 LOBs with “complete data”
 - 2,104 Companies with at least 2 Schedule P LOBs of “quality data”
 - Approx. 27,000 LOBs with at least 2 for same Company

Milliman

33

Analysis Details

- **Model Output**
 - Accident Year Totals (by Year & All Years Combined)
 - Calendar Year Totals (by Year)
 - Calendar Year Runoff Totals (by Year)
 - Ultimate Loss Ratios (by Year)
 - Incremental Results (by Year and Development Period)
 - Diagnostic Statistics

Milliman

34

Analysis Details

- **Model Options (Tests)**
 - **Test 1 – Defaults**
 - No Tail factors (i.e., 1.000)
 - BF – a priori based on hindsight L/R, **No CoV**
 - CC – Trend = 2.5%, Decay Ratio = 90%
 - **Test 2 – Selected Limiting of Incrementals**
 - **Test 3 – Selected Limiting & Suggested Heteroscedasticity Groups**

Milliman

35

Model Limitations

A Quantum Leap in Benchmarking Unpaid Claims

Model Limitations

- **Model Risk**
 - Limited to known data
 - A single model can underestimate variability
- **Systemic risk**
 - In addition to model risk
 - A shift in claims environment
- **Need to Understand Assumptions**

Major Assumption

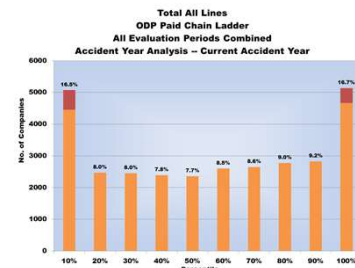
Bootstrap models (ODP & Mack) assume Chain Ladder projections are unbiased

Model Projections

Are they Unbiased?

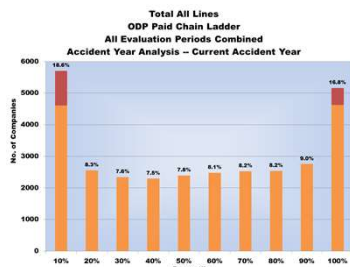
Comparison of Tests

Test 1



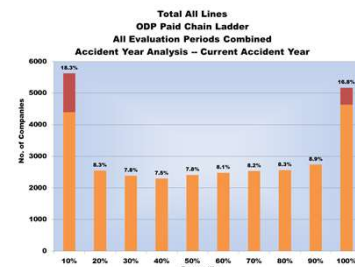
Comparison of Tests

Test 2

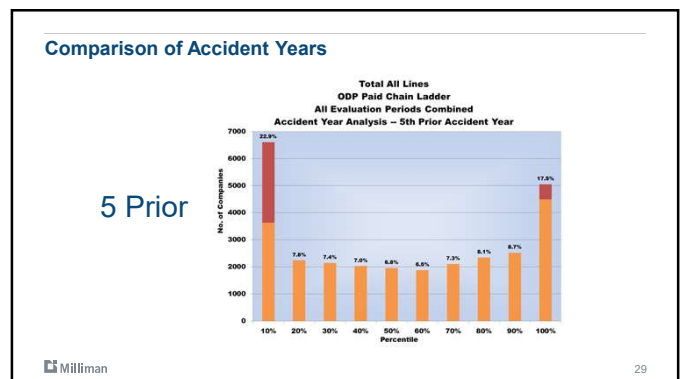
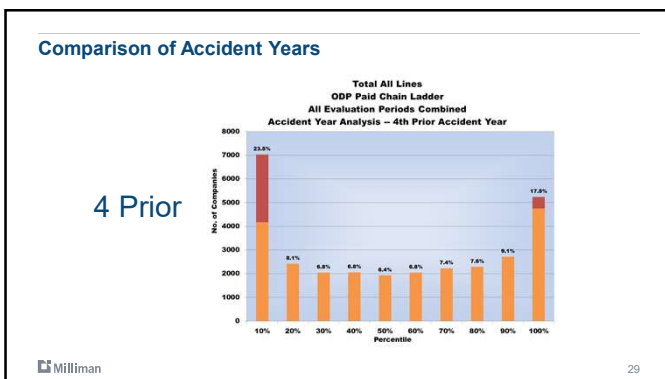
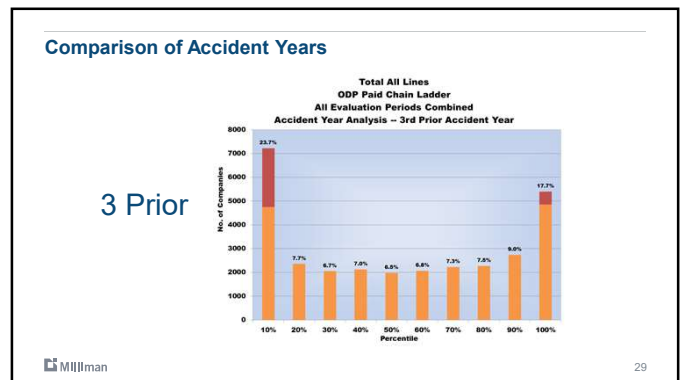
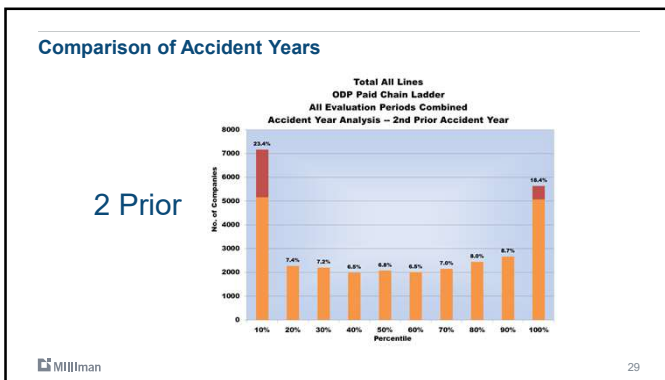
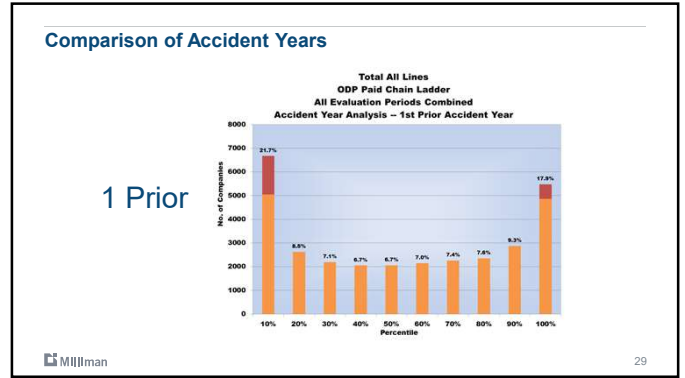
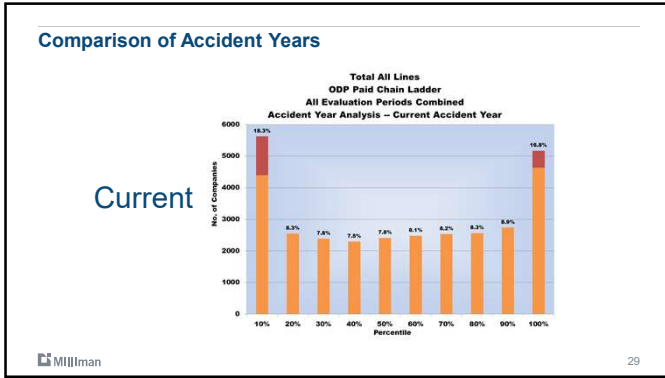


Comparison of Tests

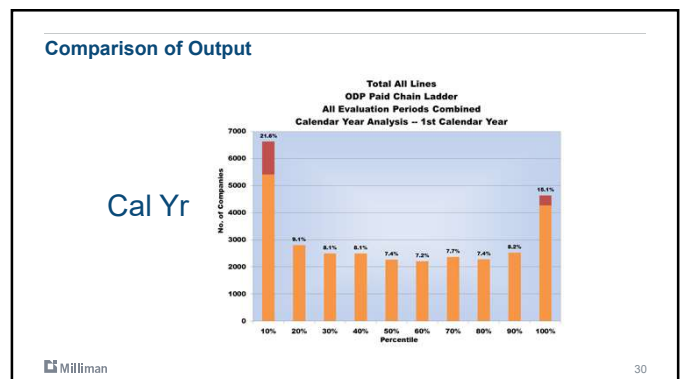
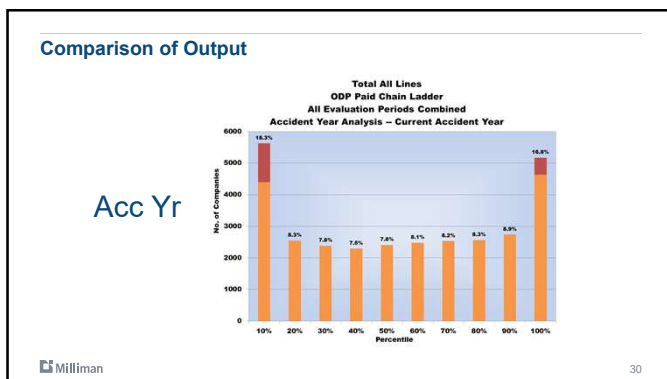
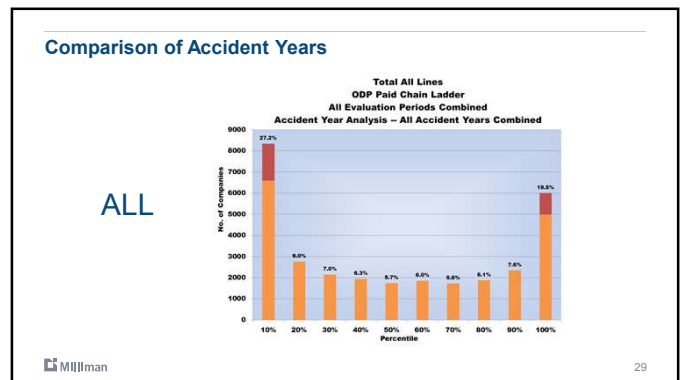
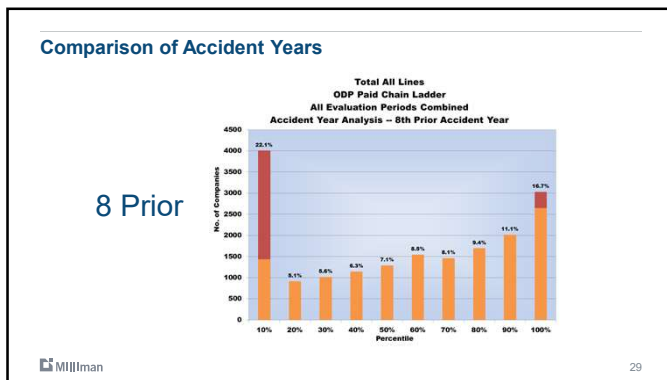
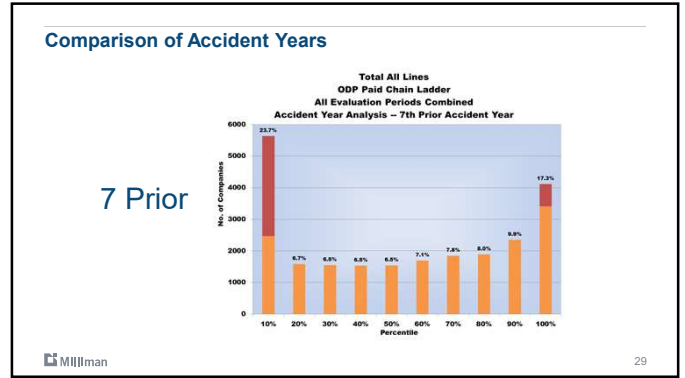
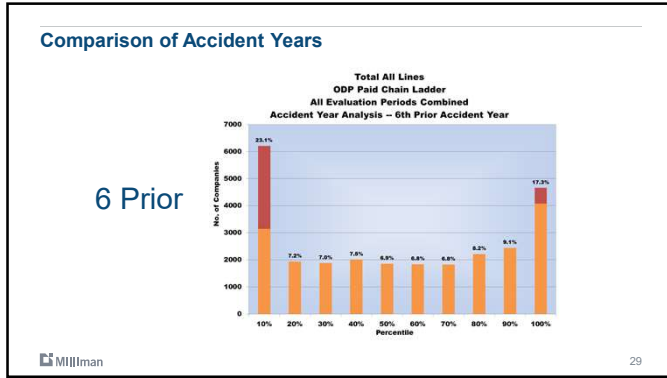
Test 3



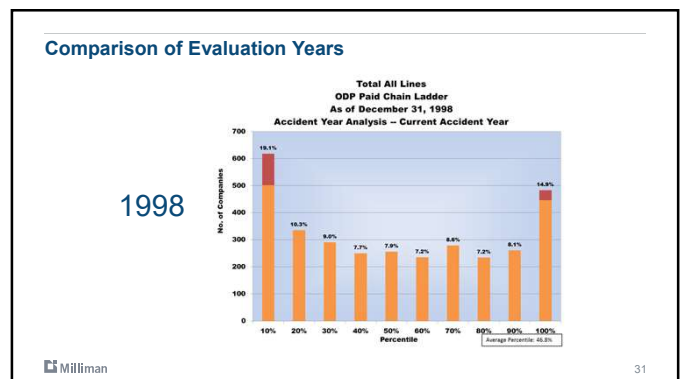
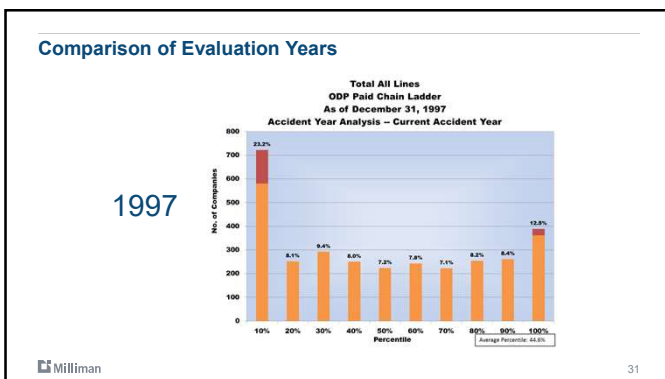
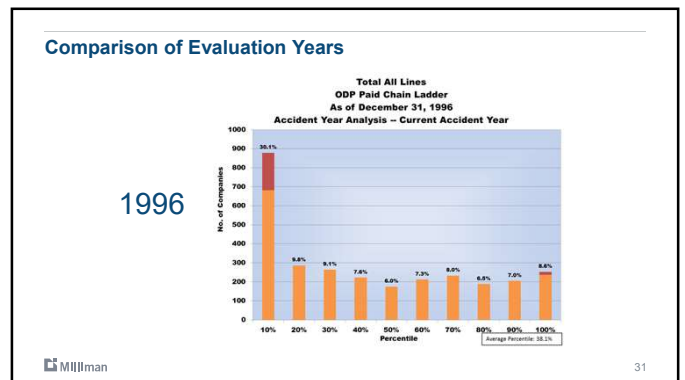
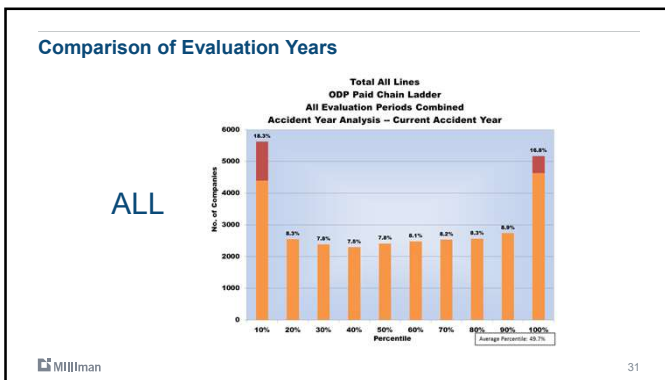
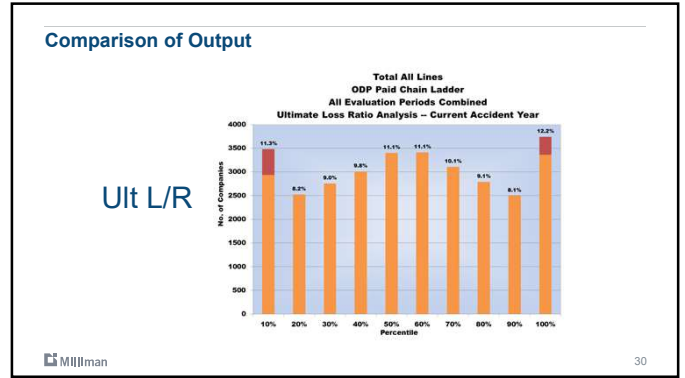
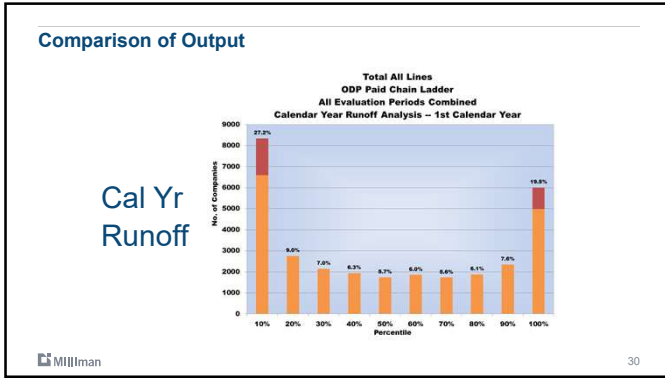
A Quantum Leap in Benchmarking Unpaid Claims



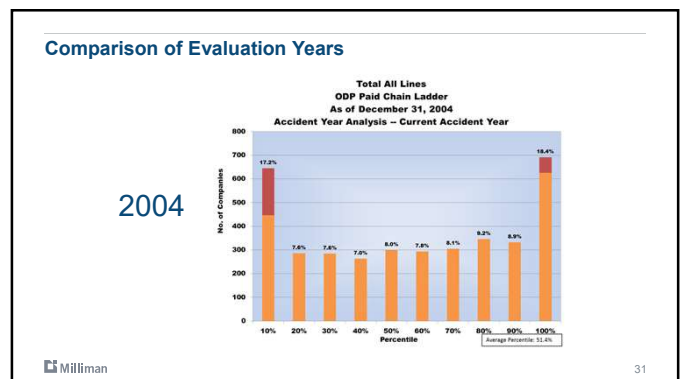
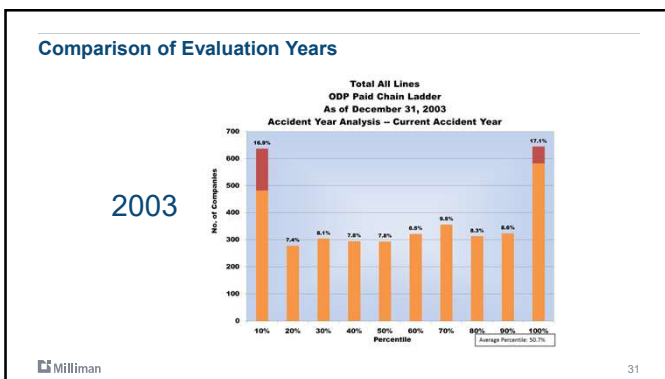
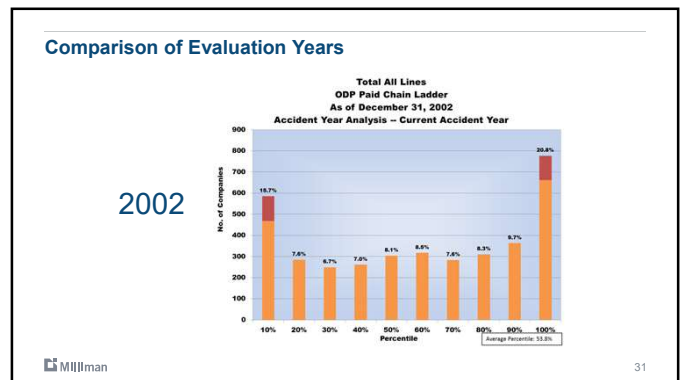
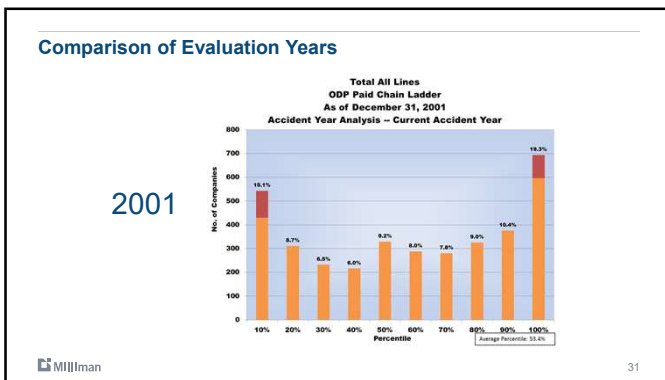
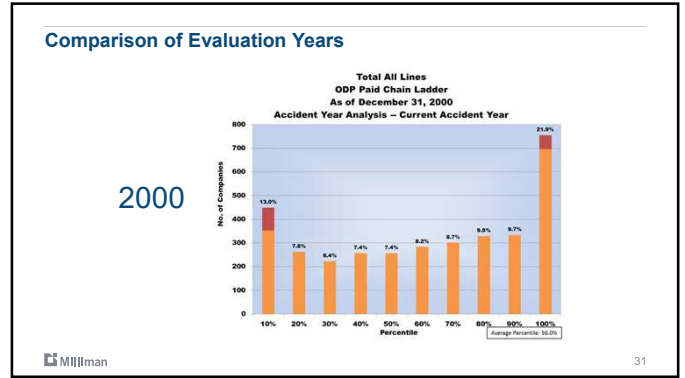
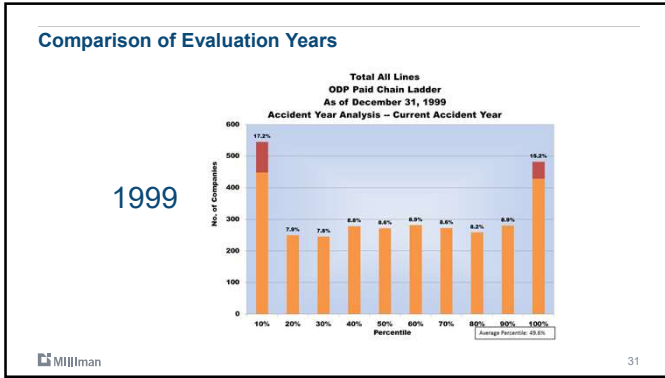
A Quantum Leap in Benchmarking Unpaid Claims



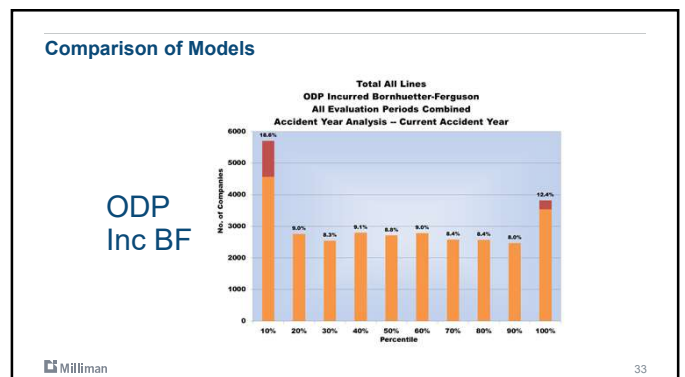
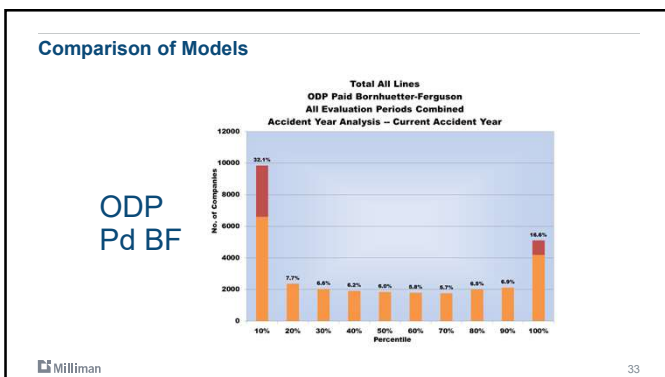
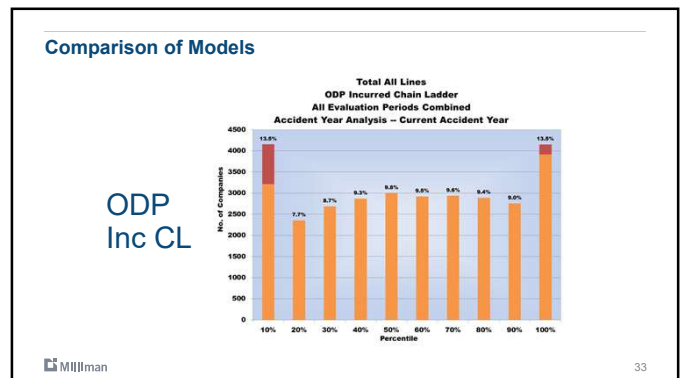
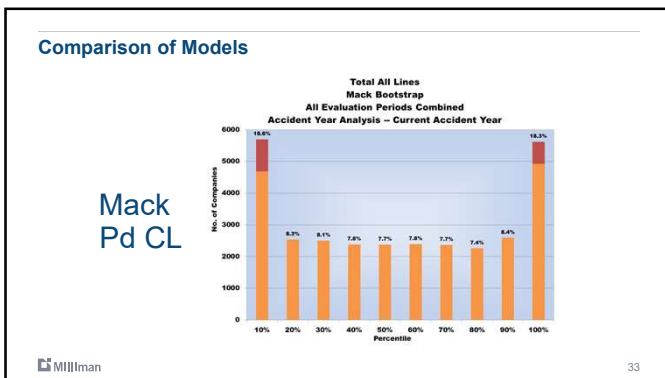
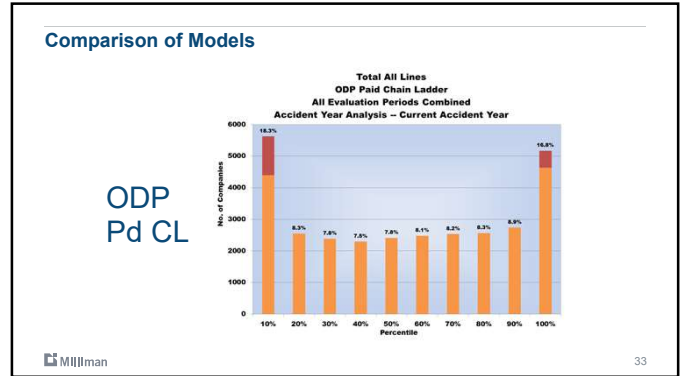
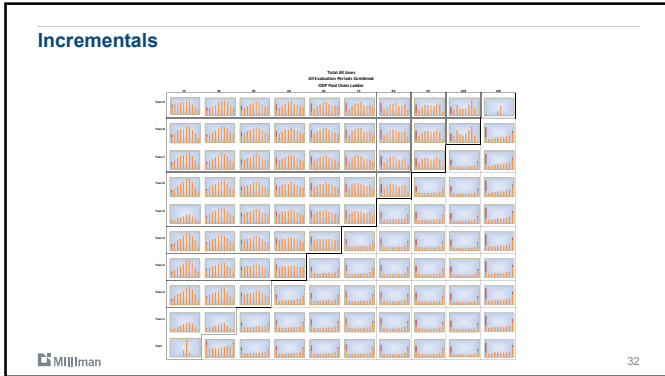
A Quantum Leap in Benchmarking Unpaid Claims



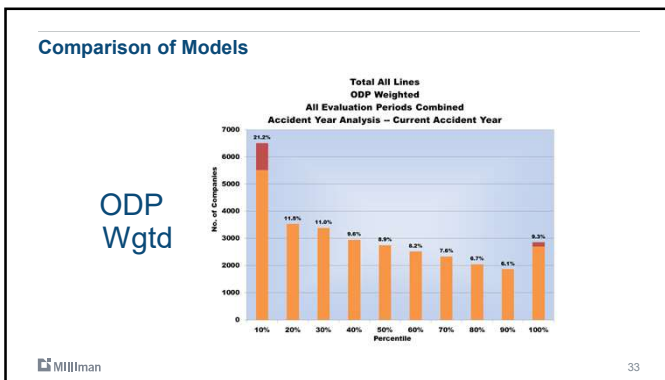
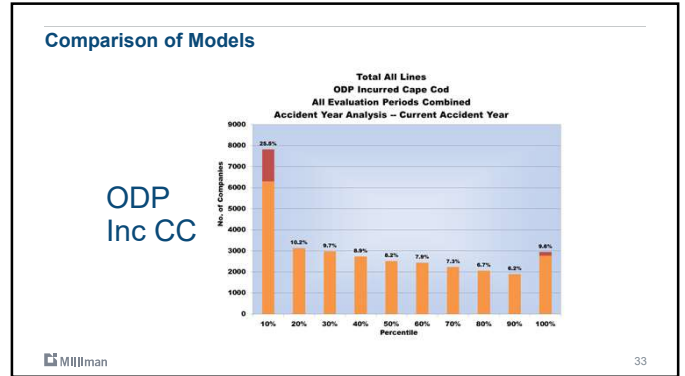
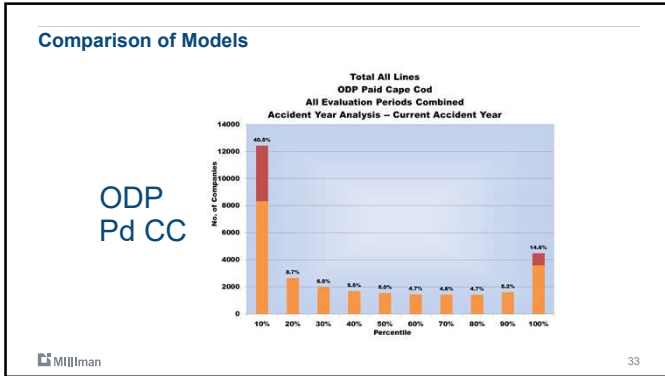
A Quantum Leap in Benchmarking Unpaid Claims



A Quantum Leap in Benchmarking Unpaid Claims



A Quantum Leap in Benchmarking Unpaid Claims



Proposed Adjustments

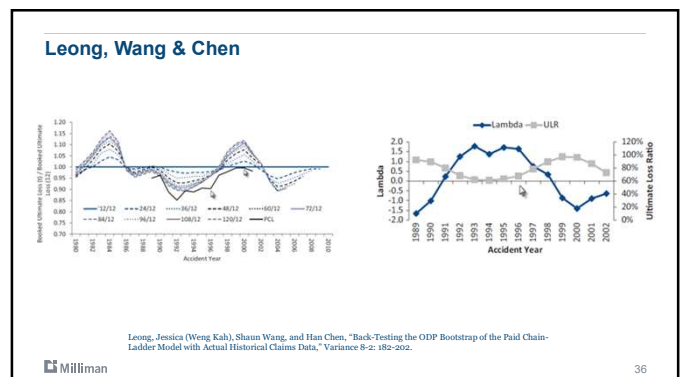
Leong, Wang & Chen

- Systemic Risk Distribution Method
 - Multiply each simulated bootstrap result by a "systemic" factor
- Wang Transform Adjustment
 - Increase the variability of the original unpaid loss distribution
 - Shift the percentiles to account for bias in methods over time
 - Relies on a parameter "Lambda" targeting an ideal histogram

**Assumes Model Risk is Systemic!
Based on Hindsight only!**

Leong, Jessica (Wong Kah), Shaun Wang, and Han Chen, "Back-Testing the ODP Bootstrap of the Paid Chain-Ladder Model with Actual Historical Claims Data," Variance 8-2: 182-202.

Milliman 35



A Quantum Leap in Benchmarking Unpaid Claims

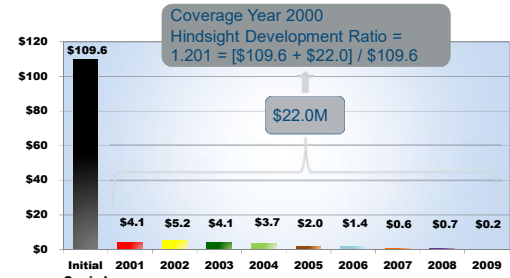
HDR Adjustment

- Shift distribution by multiplying unpaid claim estimates by the HDR
- Coefficient of variation unchanged
- Additive shift – will not address variance
- Hindsight adjustment, but we are not advocating, just testing how much bias vs. not enough variance

Milliman

37

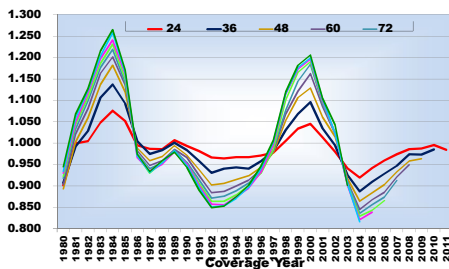
Example – Coverage Year 2000 (\$B)



Milliman

38

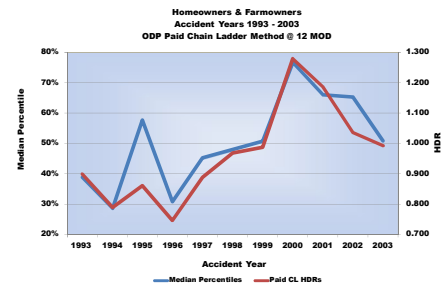
HDR by Evaluation Month



Milliman

39

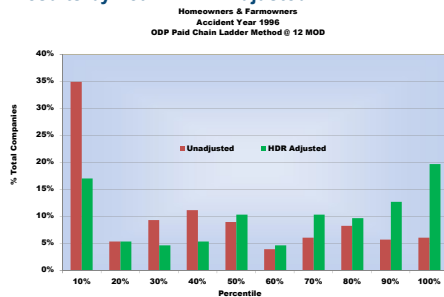
HDRs vs. Median Percentiles



Milliman

40

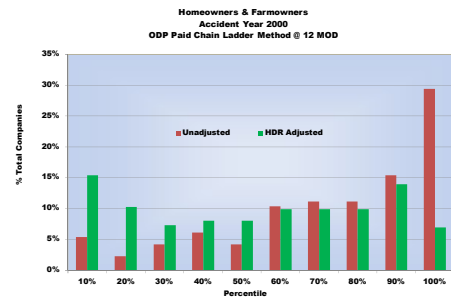
Results by Year – HDR Adjusted



Milliman

41

Results by Year – HDR Adjusted



Milliman

42

A Quantum Leap in Benchmarking Unpaid Claims



Conclusions

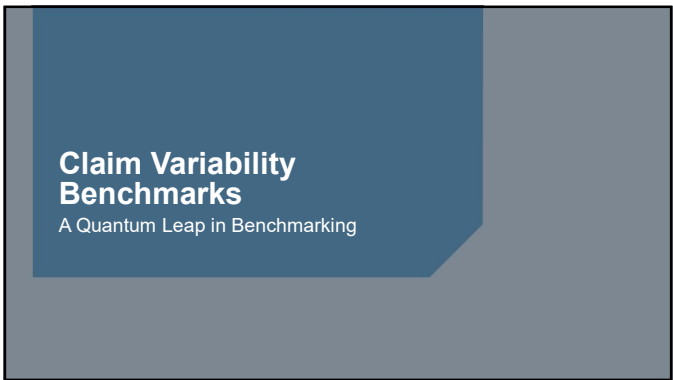
- **Goal of Ideal Histogram Unrealized by Paid CL Bootstrap**
 - Both ODP Bootstrap and Mack Bootstrap
 - Confirms Other Research
- **Other ODP Bootstraps – Much Closer to Theoretical Ideal**
 - Incurred models different (Shapland Monograph)
 - Bornhuetter-Ferguson and Cape Cod models
- **Cyclical Bias in Reserve Distributions – Paid and Incurred**
 - Consistent with Deterministic Projections

Milliman 44

Conclusions

- **“Corrections” to Other ODP Models may be Unnecessary**
- **Addressing Model Risk is very important**
 - Can’t “blindly” accept model results
 - Use diagnostics to assess model strengths / weaknesses
 - Implications for weighting
 - Still need to address systemic risks
- **Guidelines (i.e., benchmarks) to Assess Results**
 - Based on hindsight, but forward looking
 - Including Correlations
- **Distributions by LOB and Premium**

Milliman 45



Claim Variability Benchmarks
Types of Benchmarks

- 1 Loss Development Patterns
- 2 Unpaid Claim Distributions
- 3 Correlation Between Segments

Milliman 77

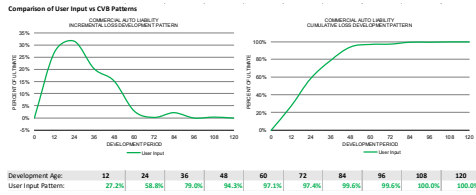
Claim Variability Benchmarks
Loss Development Patterns

- Common LDF benchmarks are “static” – one size fits all
- Back-testing includes VWA factors for all actual & simulated paid data triangles, by Schedule P Line of Business
- A “distribution” of the patterns were created for both actual and simulated data
- This allows for “dynamic” benchmarks – patterns are better tailored to your data
- You can also create a benchmark for your range of point estimates

Milliman 49

A Quantum Leap in Benchmarking Unpaid Claims

Claim Variability Benchmarks Loss Development Patterns



As an example of how you might use this information, suppose you are analyzing Commercial Auto data and have selected the following LDF pattern.



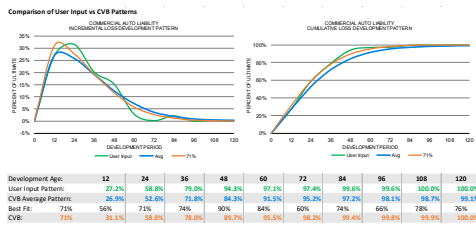
Claim Variability Benchmarks Loss Development Patterns



A typical benchmark is based on an overall average pattern, which may or may not provide a reasonable fit.



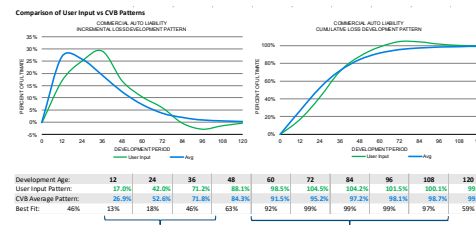
Claim Variability Benchmarks Loss Development Patterns



But using average patterns from thousands of companies, you can search percentiles for a better fit.



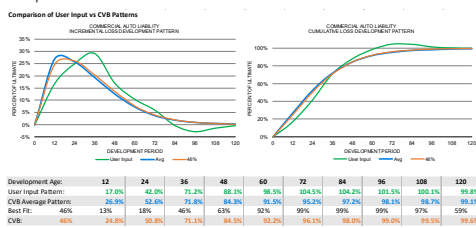
Claim Variability Benchmarks Loss Development Patterns



While a single percentile can often provide a better fit than the overall average, you might find that your pattern is slower than average in early periods and faster in later periods. Or vice versa.



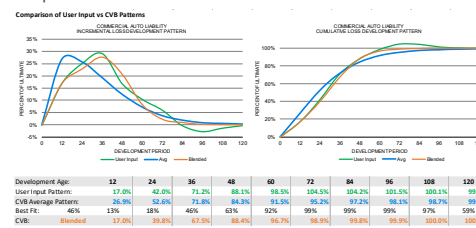
Claim Variability Benchmarks Loss Development Patterns



In this example, a single percentile is better than the average, but only marginally better.



Claim Variability Benchmarks Loss Development Patterns

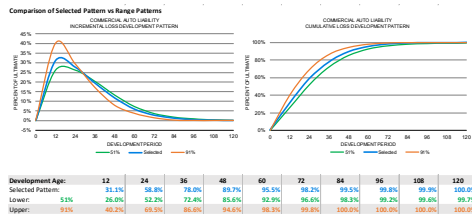


But, by blending percentiles you can create an even more customized benchmark pattern.



A Quantum Leap in Benchmarking Unpaid Claims

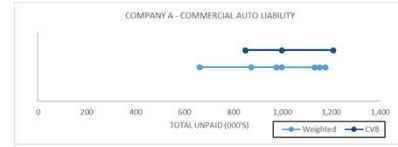
Claim Variability Benchmarks Loss Development Patterns



- To develop a range, you could calculate new unpaid claim estimates by selecting development patterns +/- X% from the best fit.

Milliman

Claim Variability Benchmarks Loss Development Patterns



- The range from the selected benchmark patterns can then be compared to the estimates from a traditional range.

Milliman

Claim Variability Benchmarks Loss Development Patterns



- The ranges for each segment can be combined into a range for the entire company.

Milliman

Claim Variability Benchmarks Types of Benchmarks

- Loss Development Patterns
- Unpaid Claim Distributions
- Correlation Between Segments

Milliman

47

Claim Variability Benchmarks Unpaid Claim Distributions

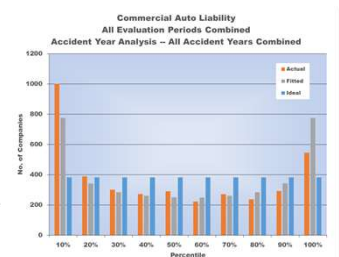
- For each Schedule P LOB, the back-testing results contain thousands of simulated distributions for companies of all different sizes
- Regression models were used to fit the distributions by premium volume for each of the Acc Yr, Cal Yr, Cal Yr Runoff, and Loss Ratio distributions
- Fitted results were smoothed to be consistent between distribution types and to conform with statistical properties – e.g., less exposure = more risk
- Algorithm allows for a variety of customizations – e.g., development patterns
- Underestimation of unpaid claim distributions can impact required capital, reinsurance, pricing, risk margins, etc.
- Overestimation is also problematic – e.g., capital does not match risk appetite

Milliman

56

Claim Variability Benchmarks Unpaid Claim Distributions

- Variance Adjustment Factors are used to correct for back-testing results
- Separate variance adjustments factors for Loss Ratio distributions
- For example, this is the Acc Yr adjustment for Commercial Auto
- “Fitted” results still appear to under-estimate, but this is reserve cycle affect



Milliman

57

A Quantum Leap in Benchmarking Unpaid Claims

Claim Variability Benchmarks Unpaid Claim Distributions

- The regression model adjusts assumptions to fit statistical properties.
- For example, consider smaller vs larger number of exposures:

Small Sample Line of Business (Using CA) Accident Year Guidelines (000's)						Large Sample Line of Business (Using CA) Accident Year Guidelines (000's)					
Acc Yr	Premium	L/R	Mean	Std Dev	CoV	Acc Yr	Premium	L/R	Mean	Std Dev	CoV
2000	5,115	75.3%	50	127	255.3%	2010	55,848	75.3%	497	991	199.6%
2001	5,302	77.1%	76	157	204.9%	2011	53,618	77.1%	766	1,015	132.8%
2012	5,427	79.4%	121	229	189.3%	2012	54,273	79.4%	1,312	1,312	108.3%
2013	5,508	81.7%	215	322	149.8%	2013	55,660	81.7%	2,100	2,662	127.8%
2014	5,668	83.5%	398	495	124.4%	2014	56,679	83.5%	3,976	2,558	64.3%
2015	5,907	82.0%	762	708	92.9%	2015	59,070	82.0%	7,625	4,029	52.8%
2016	6,277	79.2%	1,405	966	69.7%	2016	62,769	79.2%	14,048	6,512	43.8%
2017	6,780	74.9%	2,430	1,523	63.2%	2017	67,796	74.9%	24,097	9,580	38.8%
2018	7,214	73.8%	3,893	2,264	59.2%	2018	72,538	73.8%	35,920	14,503	37.3%
Total	5,121	77.4%	3,710	2,419	64.6%	Total	57,213	77.4%	13,722	7,101	52.2%

Milliman

Claim Variability Benchmarks Unpaid Claim Distributions

- The regression model allows for other customizations.
- For example, consider a faster development pattern:

Slower Sample Line of Business (Using CA) Accident Year Guidelines (000's)						Faster Sample Line of Business (Using CA) Accident Year Guidelines (000's)					
Acc Yr	Premium	L/R	Mean	Std Dev	CoV	Acc Yr	Premium	L/R	Mean	Std Dev	CoV
2000	20,409	75.3%	199	415	208.7%	2010	20,409	75.3%	61	147	242.9%
2001	21,207	77.1%	306	442	144.7%	2011	21,207	77.1%	87	168	193.9%
2012	21,709	79.4%	485	590	121.7%	2012	21,709	79.4%	152	258	169.5%
2013	22,022	81.7%	860	708	81.4%	2013	22,022	81.7%	376	411	104.7%
2014	22,671	82.5%	1,590	1,183	74.4%	2014	22,671	82.5%	895	780	87.1%
2015	23,628	82.0%	3,050	1,815	59.5%	2015	23,628	82.0%	2,046	1,328	64.9%
2016	25,108	79.2%	5,619	3,095	45.0%	2016	25,108	79.2%	4,421	2,203	49.8%
2017	27,118	74.9%	9,639	4,210	43.7%	2017	27,118	74.9%	8,415	3,754	44.6%
2018	28,865	73.8%	15,572	6,345	40.7%	2018	28,865	73.8%	14,580	6,033	41.1%
Total	215,920	77.4%	7,812	4,918	62.1%	Total	215,920	77.4%	11,173	7,140	63.6%

Milliman

Claim Variability Benchmarks Unpaid Claim Distributions

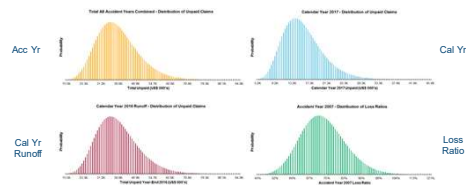
- The regression model includes four different types of results:

Sample Line of Business (Using CA) Accident Year Guidelines (000's)						Sample Line of Business (Using CA) Calendar Year Guidelines (000's)					
Acc Yr	Premium	L/R	Mean	Std Dev	CoV	Acc Yr	Premium	L/R	Mean	Std Dev	CoV
2000	20,409	75.3%	199	415	208.7%	2010	212,568	77.0%	4,621	4,402	30.1%
2001	21,207	77.1%	306	442	144.7%	2011	202,786	78.2%	3,628	3,441	30.7%
2012	21,709	79.4%	485	590	121.7%	2012	192,229	78.0%	2,780	2,594	46.1%
2013	22,022	81.7%	860	708	81.4%	2013	191,422	78.8%	3,541	3,495	58.5%
2014	22,671	82.5%	1,590	1,183	74.4%	2014	189,622	79.7%	2,728	2,292	36.7%
2015	23,628	82.0%	3,050	1,815	59.5%	2015	187,389	78.2%	910	904	38.8%
2016	25,108	79.2%	5,619	3,095	45.0%	2016	186,909	77.2%	970	792	122.8%
2017	27,118	74.9%	9,639	4,210	43.7%	2017	185,881	75.9%	342	562	147.9%
2018	28,865	73.8%	15,572	6,345	40.7%	2018	175,678	76.2%	452	222	202.0%
Total	212,528	77.4%	17,812	8,264	46.0%	Total	17,213	81.0%	6,704	24.0%	

Milliman

Claim Variability Benchmarks Unpaid Claim Distributions

- In Excel, these are easy to graph:

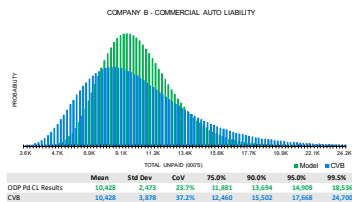


Milliman

59

Claim Variability Benchmarks Unpaid Claim Distributions

- Compared to "single" model approach, the typical estimate has less variance than the benchmark:

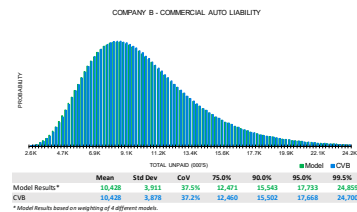


Milliman

60

Claim Variability Benchmarks Unpaid Claim Distributions

- Compared to "multiple" model approach, the typical estimate closer to the benchmark:



Milliman

61

A Quantum Leap in Benchmarking Unpaid Claims

Claim Variability Benchmarks

Types of Benchmarks

- 1 Loss Development Patterns
- 2 Unpaid Claim Distributions
- 3 Correlation Between Segments

Milliman 47

Claim Variability Benchmarks

Correlation Between Segments

- Back-testing output includes correlation statistics between all pairs of LOBs within a company (i.e., if there was more than one 'complete' LOB)
- Output includes both paid and incurred, before and after optimal hetero adjustments
- The mean and std dev (unweighted and weighted) for all specific pairs (i.e., between two specific LOBs) was measured
- Weights based on 1 minus P-Value, since the lower the P-Value the more statistically significant the correlation
- Industry benchmarks have long been needed

Milliman 67

Claim Variability Benchmarks

Correlation Between Segments

- For example, consider the weighted results for 4 LOBs using 1996 data:

Means					Means					Standard Deviations				
COMPANY A Model Correlation					COMPANY A CVB Correlation - Means					COMPANY A CVB Correlation - Std Dev				
	CA	MPL-O	PL-O	WC		CA	MPL-O	PL-O	WC		CA	MPL-O	PL-O	WC
CA	100%	34.1%	-17.4%	46.6%	CA	100%	15.8%	12.0%	14.1%	CA		21.9%	24.7%	27.6%
MPL-O	34.1%	100%	-19.4%	29.9%	MPL-O	15.8%	100%	10.3%	-3.4%	MPL-O	21.9%	100%	20.2%	21.5%
PL-O	-17.4%	-19.4%	100%	10.5%	PL-O	12.0%	10.3%	100%	11.6%	PL-O	24.7%	20.2%	100%	24.6%
WC	46.6%	29.9%	10.5%	100%	WC	14.1%	-3.4%	11.6%	100%	WC	27.6%	21.5%	24.6%	100%

Modelled Correlation | Correlation Benchmarks

Milliman 63

Claim Variability Benchmarks

Correlation Between Segments

- Consistent with individual segments, aggregates using a "single" model approach tend to be narrower than benchmarks:

	Mean	Std Dev	CV	75.0%	90.0%	95.0%
Model Results*	80,139	35,104	13.1%	65,732	91,940	113,233
CVB	80,139	20,222	25.2%	59,960	100,000	147,377
TVAR Estimates				94,837	100,000	104,133
Model Results*				107,438	120,822	130,133
Capital Required	13,658	15,922	24,994	35,883		
Model Results*				27,261	40,093	50,161
CVB						79,096

*Using top 5% GDP Benchmark model for PAID data for each LOB

Milliman 64

Claim Variability Benchmarks

Correlation Between Segments

- Consistent with individual segments, aggregates using a "multiple" model approach tend to be closer to benchmarks:

	Mean	Std Dev	CV	75.0%	90.0%	95.0%
Model Results*	80,139	35,104	13.1%	65,732	91,940	113,233
CVB	80,139	20,222	25.2%	59,960	100,000	147,377
TVAR Estimates				94,837	100,000	104,133
Model Results*	113,720	131,333	143,075	184,053		
CVB	105,420	130,843	130,200	178,692		
Capital Required	13,658	15,924	24,994	35,884		
Model Results*				27,261	40,093	50,161
CVB						79,096

*Using Results based on weighting of 4 different models for each LOB

Milliman 65

Claim Variability Benchmarks

Other Potential Uses

- Calculating average durations for future cash flows
- Calculating reserve risk margins based on the expected unpaid claim runoff – e.g., Solvency II or IFRS-17
- Assessing the variance parameter for a priori loss ratio assumptions in models
- Creating back-testing benchmarks for ERM thresholds
- Other uses which are only limited by your imagination

Milliman 73

A Quantum Leap in Benchmarking Unpaid Claims



Milliman
IT TAKES VISION

Any Final Questions?

Mark R. Shapland, FCAS, FSA, MAAA
Liberty House, Unit 501, Level 5
DIFC P.O. Box 506784
Dubai, United Arab Emirates
Tel: +971 4 386 6990
Mobile: +971 56 179 1532
mark.shapland@milliman.com