Cluster Modeling Overview

Liabilities and Assets

Presented by:
Craig Reynolds, FSA, MAAA
Consulting Actuary

November 13, 2008
Actuaries Club of the Southwest

Milliman

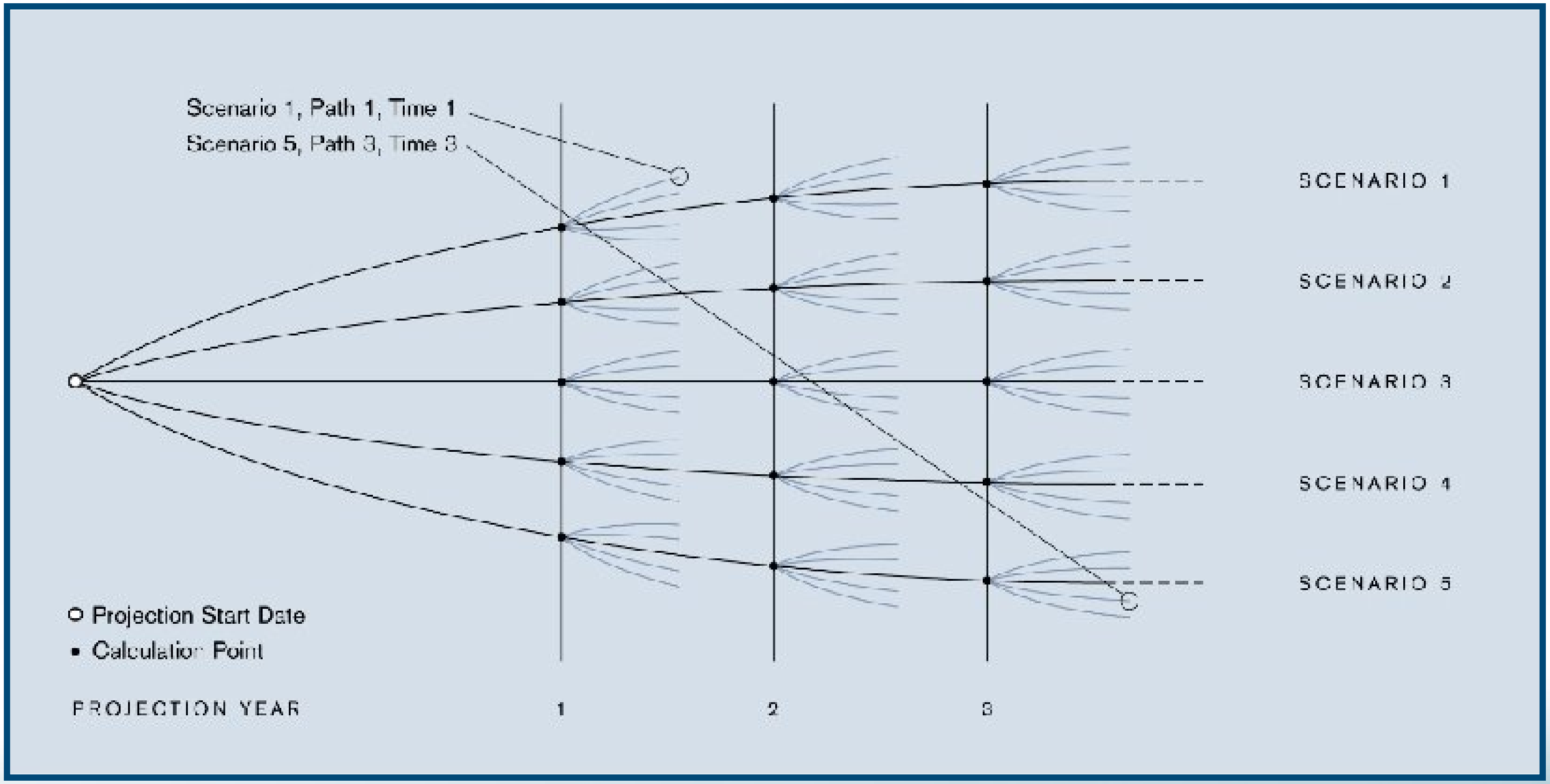
Trends in Actuarial Modeling

Past	Present	Future
Actuarial modeling required for calculations to run in an acceptable timeframe	Faster hardware and software often make seriatim calculations practical •Grid processing •Job threading •Faster processors	**Cluster Modeling makes nested stochastic and massive stochastic runs practical**

Milliman

Modeling May Always be Necessary

- Particularly nested stochastics (stochastic in stochastic)
- Examples: IFRS, FAS 133, SOP 03-1, Dynamic Hedging, PBA, Option Pricing, VA CARVM, C-3 Phase 2



Milliman

Nested Stochastic Runtimes

- Sample calculation specifications
  - 1 million policies
  - 30-year projections
  - Quarterly calculations of IFRS or other stochastic reserves across 500 paths
  - 10,000 scenarios
- Implications-Sometimes seriatim cannot be done
  - 600 trillion policy-path projections
  - At 1000 cell paths per second, this is still:
    - 600 billion seconds
    - 19 thousand years
- Clearly we cannot rely on hardware or software alone!

Cluster Modeling Does it Better

- Do not ask: **To model or not to model?**
- Instead ask: **When you have to model, how to do it best?**

Living in a World With Modeling

- Classic Modeling Techniques
  - Some rule-based (age modeling, issue-date modeling)
  - Some judgment-based (minor plans to major plans)
  - Focused on validation of initial balance sheet
  - Assumes that reproduction of initial amounts implies good reproduction of future earnings
- Challenges
  - Keeping up-to-date with new plans
  - Managing and measuring model noise
  - Making auditors happy

Cluster Modeling Diagram-Two Dimensions

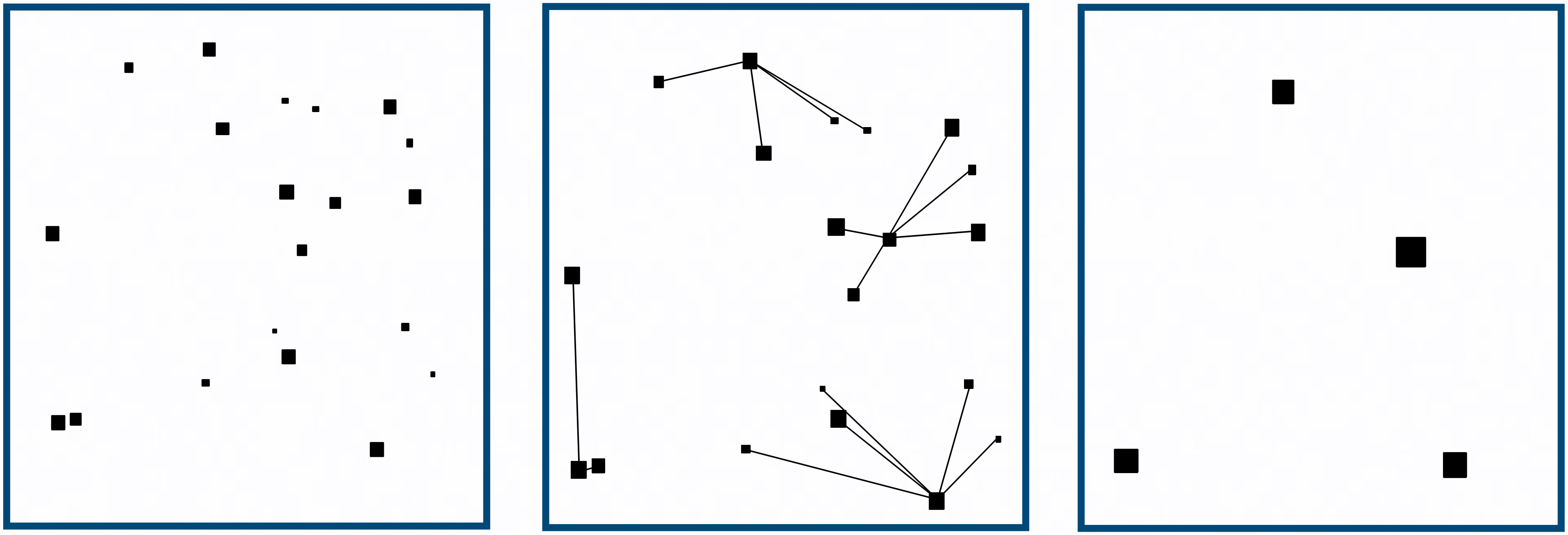
(Liability Example: Opening reserve and FY premium)

(Asset Example: Book/Par Ratio and Yield to Maturity)

Two Dimensional Plot of Policies of Various Sizes

Assign Policies to Clusters

Gross up Central Points



Cluster Modeling Eases Challenges

- Any product or asset type
- Better compression ratios for a given model-to-actual fit
- Easily automated with little upfront effort
- Maintained and applied in similar ways at later valuation dates
- Allows customization to place different priorities on different measures of model fit
- Can be applied to seriatim or modeled in-force
- Allows easy adjustment to the number of model points to produce more or less model granularity, depending on the application
- Allows easy on-the-fly analysis of model fit for differing levels of model granularity, without rerunning a model

Key Cluster Modeling Concepts

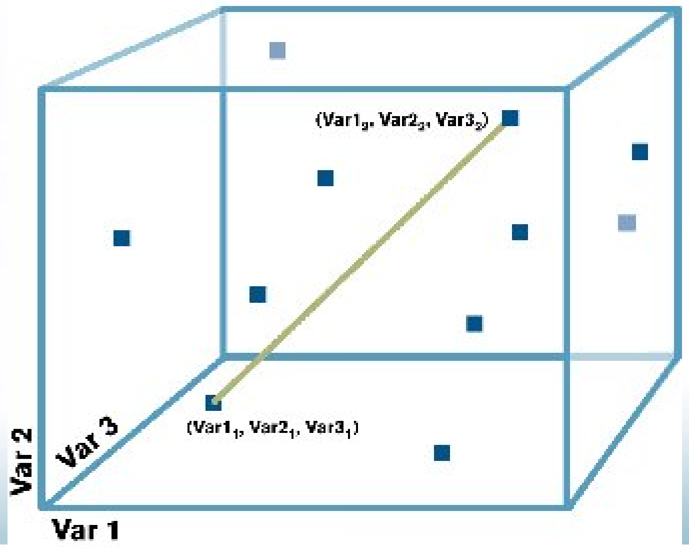
- *Location Variable*: Any value that you want the model to closely reproduce, e.g.,
  - Opening reserves or premiums in-force
  - First-year premiums
  - First-year claims
  - Net-liability cash flow in each of the first five years
  - Asset coupon rate
  - Book / Par ratio
  - Present value of profits
- Values may be normalized by dividing by sample standard deviation
- Users define the list of variables and capture their values in an inventory report

Milliman

Key Cluster Modeling Concepts

- *Distance Function*: A measure to show how “far away” any two policies or cusips are from each other in n-dimensional space
- Euclidean distance operating on normalized location-variable values, with each variable representing one spatial dimension
- May assign weights to scale up or down distances in certain dimensions to be consistent with importance of this dimension

$$\sqrt{(Var1_1 - Var1_2)^2 + (Var2_1 - Var2_2)^2 + (Var3_1 - Var3_2)^2}$$



Milliman

Key Cluster Modeling Concepts

- *Size*: One component of the *importance* of each policy
  – Typically face amount or units in-force
  – Might also be account value in-force, annuity benefit amount, or some other user-defined quantity
- *Importance = (Size) * (Distance to nearest neighbor)*

Milliman

Key Cluster Modeling Concepts

- *Segment*: A group that each policy belongs in, such that no policy will be mapped outside of its group
- LOB or asset class will always be a segment
- Can also be things like premium period, insurance period, GAAP era, reserve basis, issue year, or plan code
- Use of segments shrinks compression time and may improve model mapping results across other scenarios

Cluster Modeling Algorithm

- Compute the distance of every policy from every other in its segment
- Compute the *Importance* of each policy as the product of (size) * (distance to nearest neighbor) for each policy.
- Identify the policy with the least importance. Map it to its nearest neighbor within the same segment.
- Repeat until the desired number of cells is obtained
- For each resulting cluster, pick the point in the cluster that is closest to the average location of all cells in that cluster. Use this point to represent the cluster.
- Gross up or add up all in-force data associated with the destination cell
- Review model fit
- Refine location variables and weights as desired and repeat

Case Study 1: A Life / Health Model

- 120,000 model points in original model
- Mix of traditional life and health products
- 200 model points in cluster “model of model”
- Liability focused—but could just as easily have been assets

Case Study 1: Location Variables

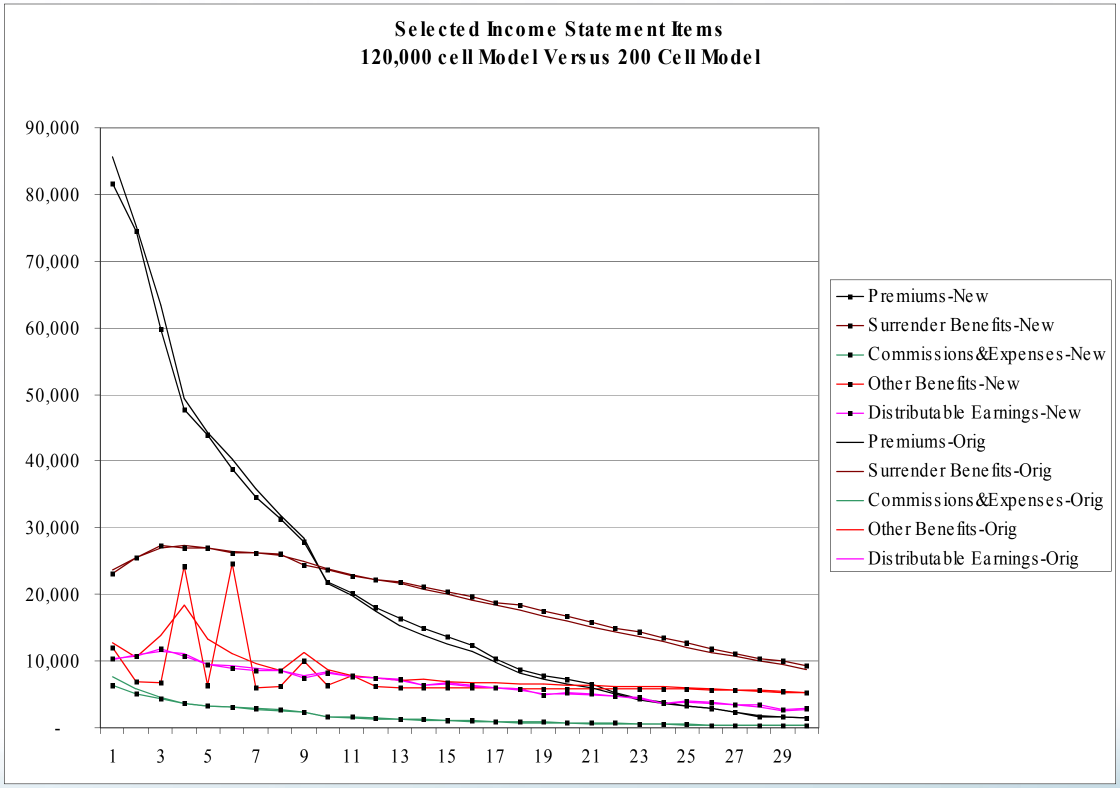
- Initial reserve (weight 1)
- First projection year premiums (weight 1)
- First projection year claims (weight 1)
- PV of proxy profits (weight 8)
- PV of proxy profits through 10 projection years (weight 6)
- PV of proxy profits through 20 projection years (weight 6)

Case Study 1: Results

	Original	New	Difference	Ratio
Initial Reserve	372,911	371,605	(1,306)	99.65%
First-Year Premiums	85,708	81,645	(4,063)	95.26%
First-Year Claims	36,485	35,162	(1,322)	96.38%
PV of Profits	154,467	154,444	(23)	99.99%
PV of Profits—10 years	77,808	77,634	(174)	99.78%
PV of Profits—20 years	119,924	120,001	77	100.06%

Case Study 1: More Results

- Excellent match on profit and most income statement items
- Limited noise is related to timing of maturity benefits—with no material bottom line impact



Milliman

Case Study 2: A Large Seriatim Term Model

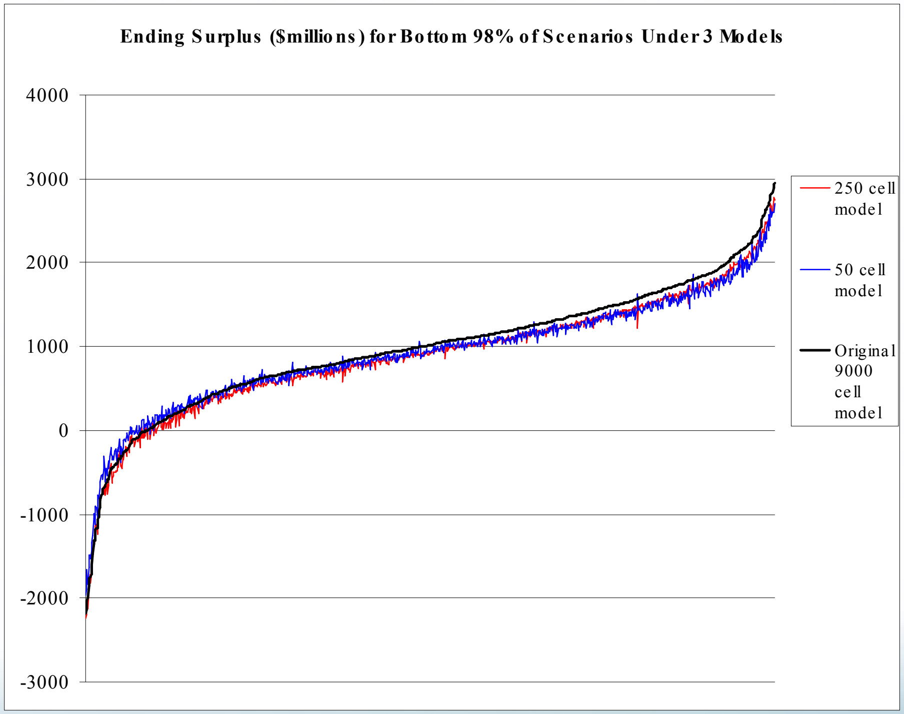
- Only the base scenario is used for calibration
- Despite this, we have excellent model fit for other scenarios with 4000 to 1 compression!

Scenario	Seriatim	300 Cell Model	Difference	Ratio
Base	4,309	4,295	14	100.3%
Mortality*115%	3,649	3,651	(3)	99.9%
Mortality*85%	4,978	4,945	33	100.7%
Lapse*115%	3,714	3,685	29	100.8%
Lapse*85%	5,251	5,266	(15)	99.7%

Milliman

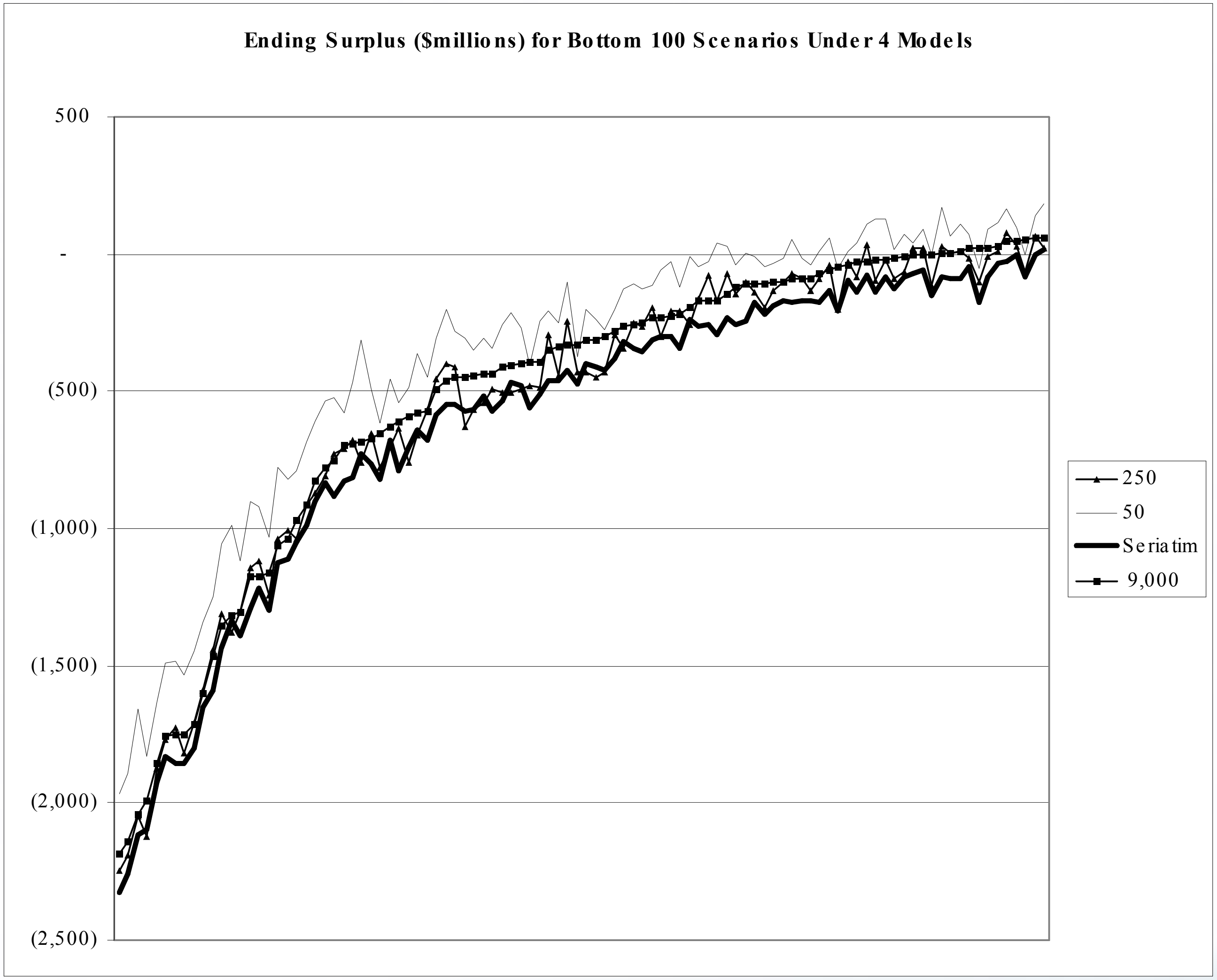
Case Study 3: A Variable Annuity Model

- 200,000 policies with GMDB, GMAB, GMWB, GMIB
- Original company classic model was 9,000 cells
- Excellent fit of cluster model to original model across scenarios, despite using only two scenarios for calibration



Milliman

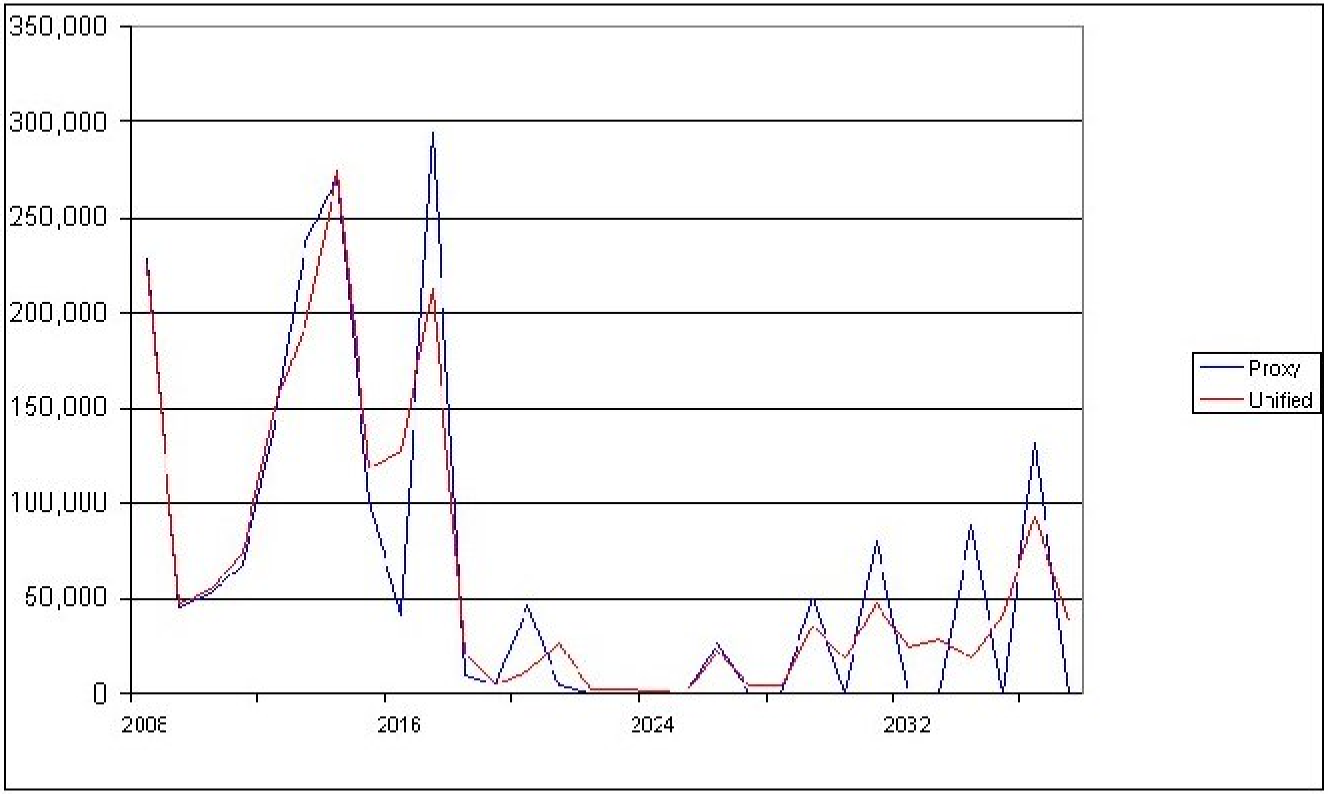
Good Fit For Tail Analysis as Well



Milliman

Case Study 4: An Existing Asset Model

- First attempt. 8-1 compression on runoff MBS and Bonds



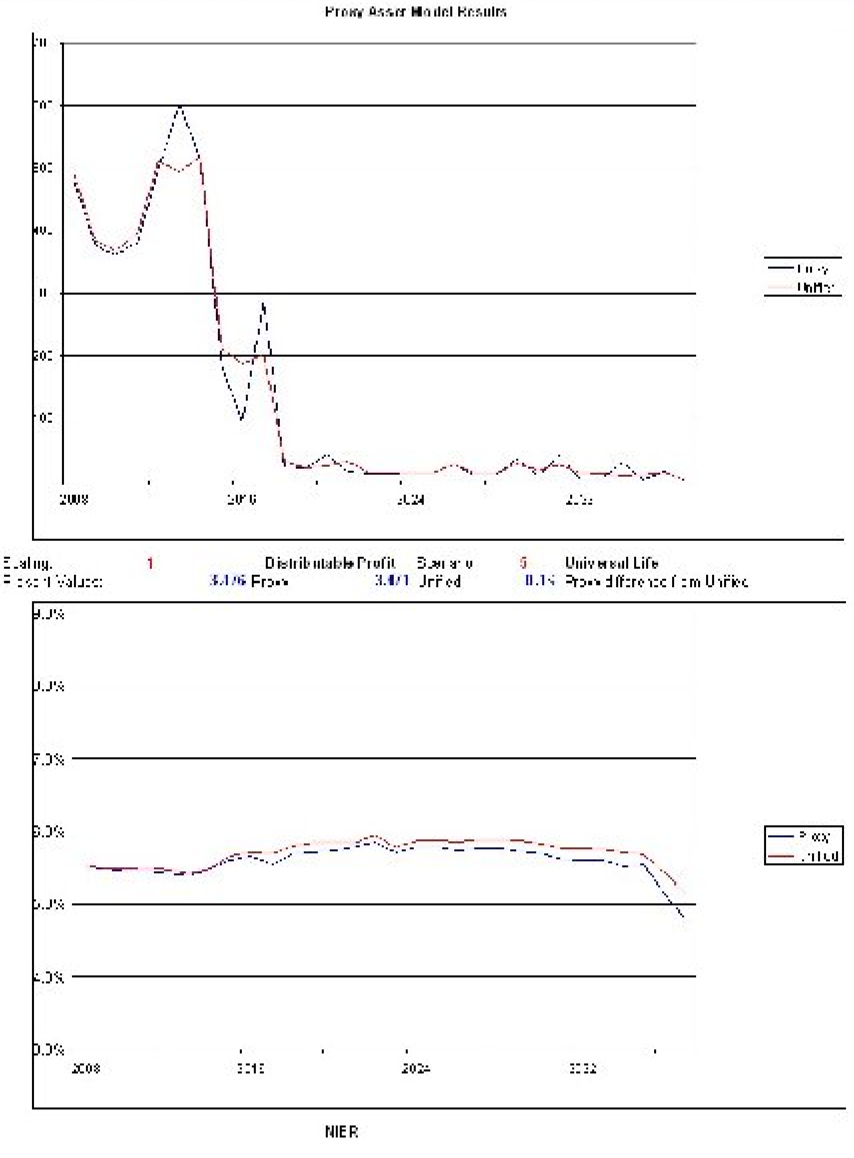
Scaling: 1,000 Principal Cash Flows Scenario 5 EEV-type projection: 30 yrs, annual

Milliman

Case Study 4: An Existing Asset Model

Proxy Asset Model
March 2008
Universal Life
1st Proxy Inforce

EEV08Q1 Stochastic Scenario	Present Value of Distributable Profits Proxy	Present Value of Distributable Profits Unified	Proxy diff from Unified	SSE as % of PVDP
1	3,429	3,423	0.2%	0.7%
2	3,359	3,359	0.0%	0.7%
3	3,120	3,127	-0.2%	1.5%
4	3,616	3,595	0.6%	2.8%
5	3,476	3,471	0.1%	0.7%
6	3,310	3,308	0.1%	0.7%
7	3,545	3,541	0.1%	0.8%
98	3,619	3,602	0.5%	2.9%
99	2,730	2,752	-0.8%	1.0%
100	3,783	3,767	0.4%	2.3%
Averages	3,356	3,351	0.1%	1.2%
Max	3,813	3,799	1.1%	3.1%
Min	2,619	2,629	-0.8%	0.6%



Next Key Development Steps

- Further validation
- Testing to determine optimal location variables and weights
- Extension to use for scenario reduction as well

Implementation Steps

- Define location variables, calibration scenarios, and inventory reports
- Identify target number of cells and assign weights to calibration variables
- Identify validation criteria
- Implement compression
- Validate
- Refine as needed
- Works with MG-ALFA® or with data from other systems