Cluster Modeling
A Practical Model and Scenario Reduction Technique

Presented by:
Wing F Wong, FSA, MAAA
Consulting Actuary
wing.wong@milliman.com

July 2009
Agenda

1. The Age of Stochastic Models
2. Cluster Modeling Concepts
3. Case Studies: Does It Work?
4. Applying Cluster Modeling to Scenario Reduction
5. Cluster Modeling versus Replicated Portfolio
The Age of Stochastic Models
Trends in Actuarial Modeling

**Past**
- **Grouping**
  - required to run in an acceptable timeframe

**Present**
- **Faster Hardware/Software**
  - often make seriatim calculations practical, along with
    - Job threading
    - Computing grid

**Future**
- **Cluster Modeling**
  - makes nested stochastic and massive stochastic runs practical
Driving Forces

- Stochastic cash flow testing/solvency test
- Principle-based reserving and risk-based capital in the US
  - Computing the CTE
- MCEV/IFRS 4 Phase 2/Solvency II in Europe
  - Fair value of options and guarantees
- Pricing of variable annuity guarantees
  - Cost of options
Credit Crisis Reminds US

- Interest rate, credit spreads, equity market – volatile

- Formula-based reserve and capital rules were designed in the stable market environment.

- Faster pace to principle-based reserve and capital, stochastic ALM.

- You will be asked to master stochastic modeling - fast.
The Need for Nested Stochastic Projections
Cluster Modeling Concepts
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!
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 Does it Better

- Do not ask: To model or not to model?
- Instead ask: When you have to model, how to do it best?
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 MG-ALFA® inventory report
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.

\[ \sqrt{(Var_{1_1} - Var_{1_2})^2 + (Var_{2_1} - Var_{2_2})^2 + (Var_{3_1} - Var_{3_2})^2} \]
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)*
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, 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 \textit{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 Studies
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

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

Selected Income Statement Items
120,000 cell Model Versus 200 Cell Model

- Surrender Benefits
- Premiums
- Dist. Earnings
- Comm & Exp
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!

<table>
<thead>
<tr>
<th>Scenario</th>
<th>1,200,000 Seriatim</th>
<th>300 Cell Model</th>
<th>Difference</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>4,309</td>
<td>4,295</td>
<td>14</td>
<td>100.3%</td>
</tr>
<tr>
<td>Mortality*115%</td>
<td>3,649</td>
<td>3,651</td>
<td>(3)</td>
<td>99.9%</td>
</tr>
<tr>
<td>Mortality*85%</td>
<td>4,978</td>
<td>4,945</td>
<td>33</td>
<td>100.7%</td>
</tr>
<tr>
<td>Lapse*115%</td>
<td>3,714</td>
<td>3,685</td>
<td>29</td>
<td>100.8%</td>
</tr>
<tr>
<td>Lapse*85%</td>
<td>5,251</td>
<td>5,266</td>
<td>(15)</td>
<td>99.7%</td>
</tr>
</tbody>
</table>
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 three scenarios for calibration
Case Study 3: A Variable Annuity Model

Ending Surplus ($millions) for 1000 Scenarios Under 3 Models

Original 9000 cells
50 cell
250 cell
Good Fit For Tail Analysis as Well

Ending Surplus ($millions) for Worst 100 Scenarios Under 4 Models
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
Cluster Models for Scenario Reduction
We can extend to scenarios

- Use the risk factors as location variables, e.g., interest rate, equity returns, bond returns, etc.

- Particularly useful for nested stochastic environment.
Case Study 5 – GMWB

- Run time from 1 hour → 3 minutes

<table>
<thead>
<tr>
<th>Plan</th>
<th>1000 Scenarios</th>
<th>50 Clustered Scenarios</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000001</td>
<td>66</td>
<td>66</td>
<td>99%</td>
</tr>
<tr>
<td>0000002</td>
<td>56</td>
<td>54</td>
<td>96%</td>
</tr>
<tr>
<td>0000003</td>
<td>59</td>
<td>59</td>
<td>99%</td>
</tr>
</tbody>
</table>
## Case Study 6 – GMAB

<table>
<thead>
<tr>
<th>Plan</th>
<th>1000 Scenarios</th>
<th>50 Clustered Scenarios</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>000001</td>
<td>39</td>
<td>40</td>
<td>103%</td>
</tr>
<tr>
<td>000002</td>
<td>33</td>
<td>34</td>
<td>104%</td>
</tr>
<tr>
<td>000003</td>
<td>34</td>
<td>36</td>
<td>106%</td>
</tr>
</tbody>
</table>
## Case Study 7 – GMIB

<table>
<thead>
<tr>
<th>Plan</th>
<th>1000 Scenarios</th>
<th>50 Clustered Scenarios</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000001</td>
<td>65</td>
<td>61</td>
<td>94%</td>
</tr>
<tr>
<td>0000002</td>
<td>58</td>
<td>54</td>
<td>92%</td>
</tr>
<tr>
<td>0000003</td>
<td>59</td>
<td>55</td>
<td>92%</td>
</tr>
</tbody>
</table>
## Case Study 8 – GMDB

<table>
<thead>
<tr>
<th>Plan</th>
<th>1000 Scenarios</th>
<th>50 Clustered Scenarios</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>000001</td>
<td>30</td>
<td>29</td>
<td>98%</td>
</tr>
<tr>
<td>000002</td>
<td>26</td>
<td>26</td>
<td>99%</td>
</tr>
<tr>
<td>000003</td>
<td>25</td>
<td>26</td>
<td>101%</td>
</tr>
</tbody>
</table>
Cluster Model versus Replicating Portfolio
Replicating Portfolio

- Replicating portfolio is a way to reduce runtime
- Search for a portfolio of assets and derivatives to represent the cash flows

- Advantage:
  - Reduce liability model to a small subset of assets
  - Asset valuations may be done by closed form solution in some cases.

- Useful for stochastic environment.
Replicating Portfolio

- Disadvantages
  - Require specialized knowledge of assets and derivatives
    - The work is likely taken out of the hands of the regular actuaries
  - “Lost” the feel of policies
  - “Lost” the link to every day financial reporting
  - May not handle policyholder behavior well
  - Does not reduce scenarios
  - Have to redo replicating portfolio if major changes in liability model
Thank you!