SOCIETY OF ACTUARIES

Life 2008 Spring Meeting
June 16-18, 2008

Session 62, Introduction to Modeling Efficiency and Scenario Reduction Techniques – Part 3

Moderator

Anthony Dardis, FSA, FIA, MAAA

Authors

Joshua P. Liu, FSA, MAAA

John O. Nigh, FSA, MAAA

Craig A. Roberts, FSA, MAAA

Data Mining, Survey Sampling, Minimize Spatial Distance

Joshua Liu, **FSA MAAA**

ING Financial Service

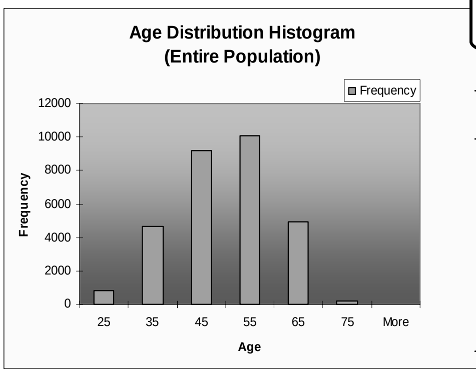
Risk Management Department/ ALM

Joshua.liu@us.ing.com

Current Challenge

- **Usually, the population is too large for the researcher to attempt to survey all of its members. A small, but carefully chosen sample can be used to represent the population. The sample reflects the characteristics of the population from which it is drawn.**

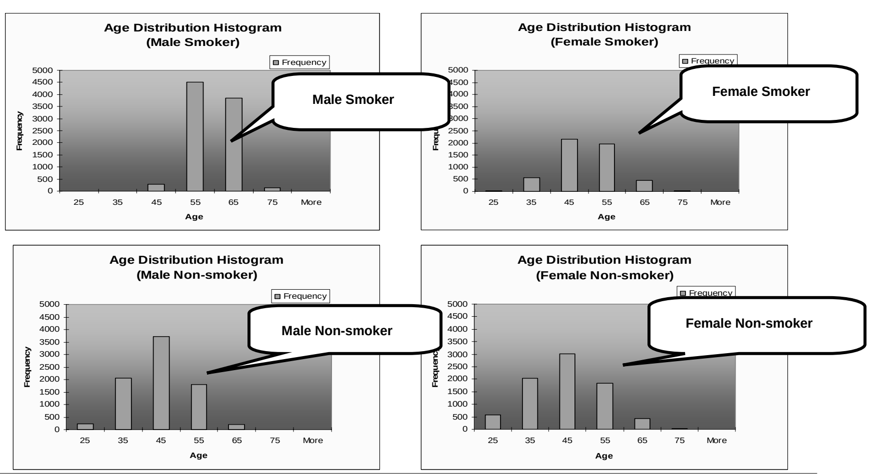
Population Data Study (Consolidation)



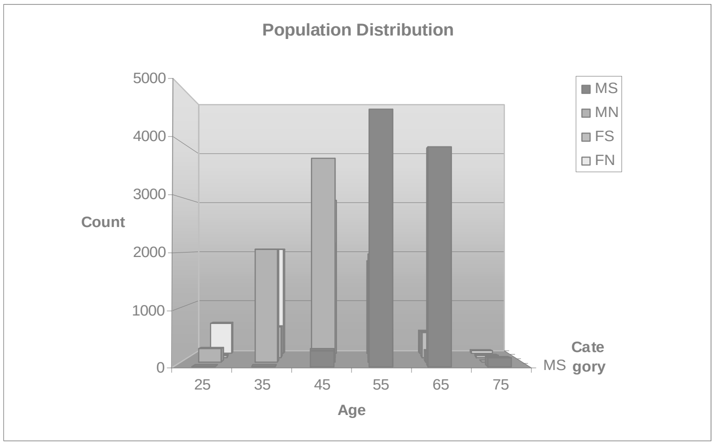
This is total population that include male smoker, male non-smoker, female smoker, and female non-smoker

Bin	*Frequency*	MS	MN	FS	FN
25	853	-	239	35	579
35	4,686	1	2,051	581	2,053
45	9,197	286	3,714	2,172	3,025
55	10,085	4,502	1,787	1,968	1,828
65	4,951	3,853	203	466	429
75	227	153	2	32	40
More	1	0	0	0	1

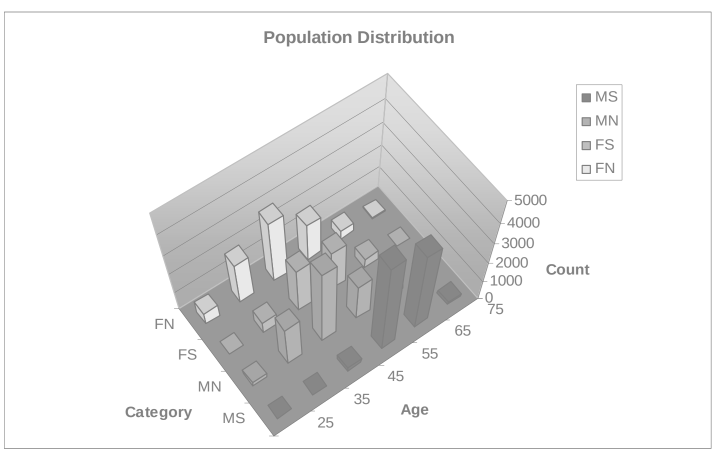
Population Data Study



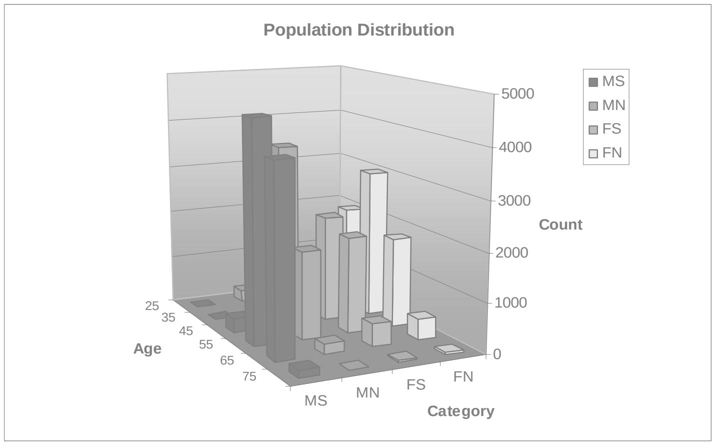
Population Data Study in 3-D



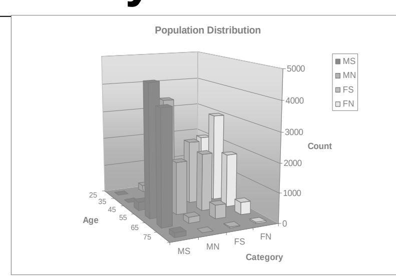
Population Data Study in 3-D (Bird-eye view)

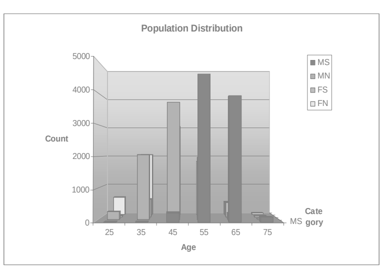


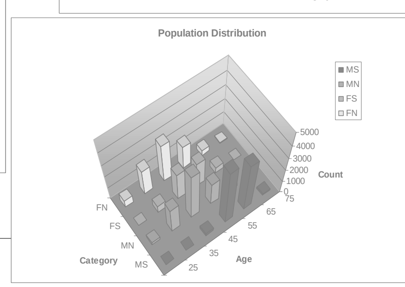
Population Data Study in 3-D (Side view)



Population Data Study







Survey Sampling

- Sampling methods are classified as either
  - Non-*probability*
  - *Probability*

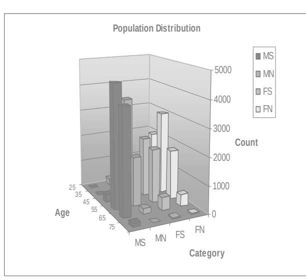
Non-probability Sampling Methods

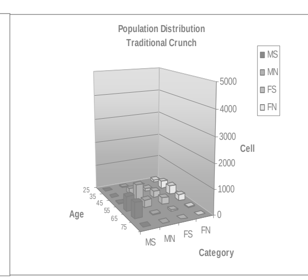
- Convenience sampling
- Judgment sampling
- Quota sampling
- Snowball sampling.

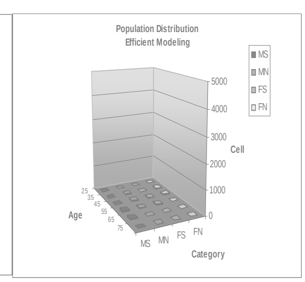
Probability Sampling Methods

- Random sampling
- Systematic sampling
- Stratified sampling
- Cluster sampling

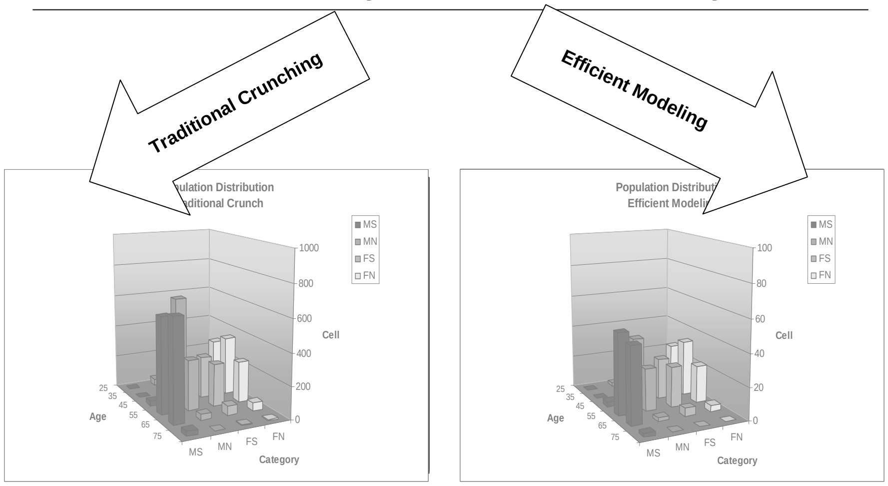
Sample Size Comparison







Traditional Crunching v.s Efficient Modeling



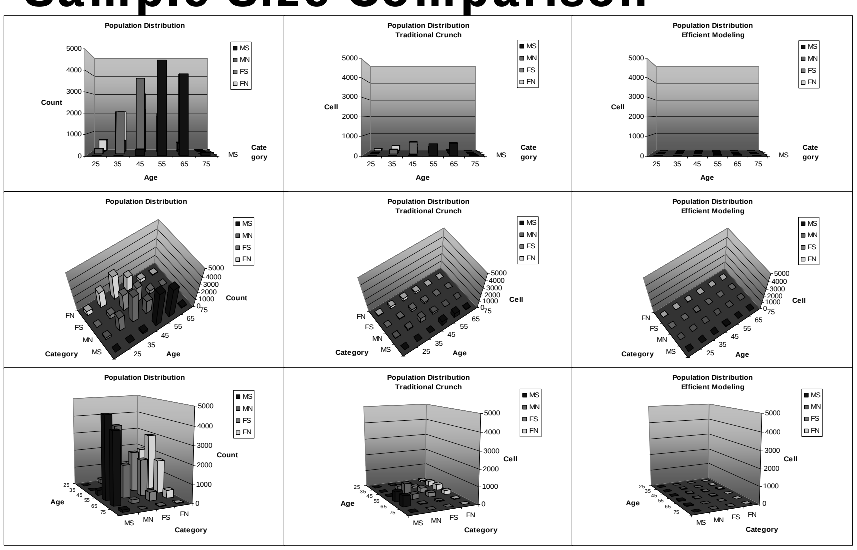
Sampling Size

- It is not straightforward to define a frame representative of the population, it is more important to understand the cause system of which the population are outcomes and to ensure that all sources of variation are embraced in the frame.

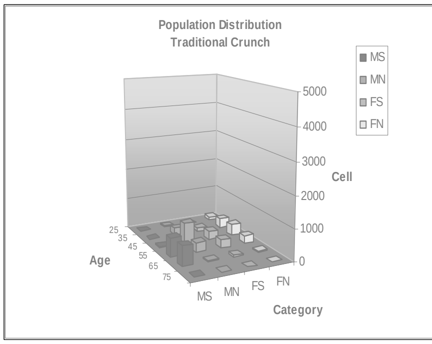
Why Stratified Sampling

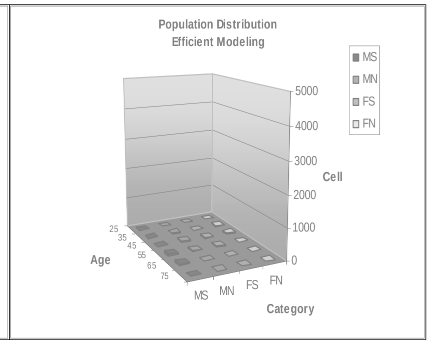
- It is superior to random sampling because it reduces sampling error.
- A stratum is a subset of the population that share at least one common characteristic.
- Examples of stratums might be
  - Males and Females
  - Smokers and Non-smokers
  - Younger and Older

Sample Size Comparison

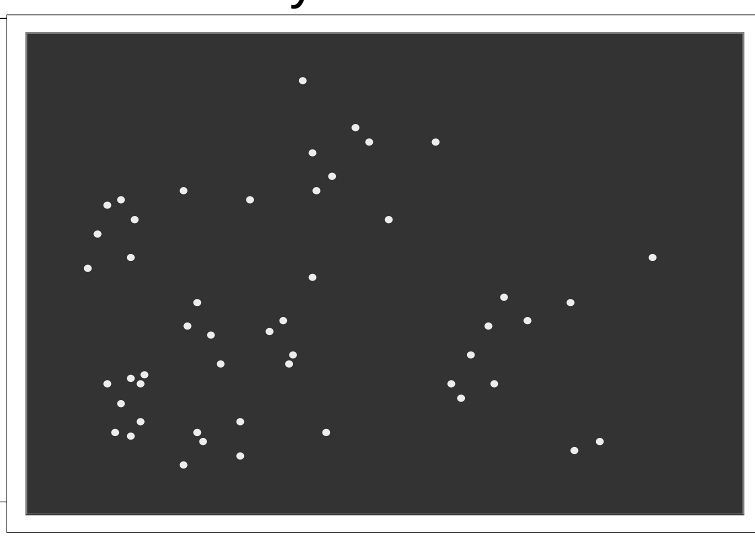


Sample Size Comparison
Traditional Crunching v.s Efficient Modeling

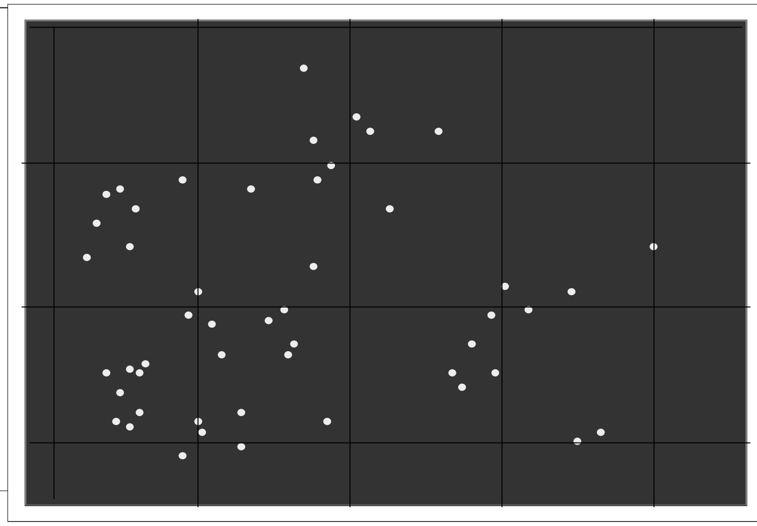




Cluster Analysis



Survey Sampling



Step of process for Stratify Sampling

- The researcher first identifies the relevant stratums and their actual representation in the population.
- Random sampling is then used to select a *sufficient* number of subjects from each stratum.
- "*Sufficient*" refers to a sample size large enough for us to be reasonably confident that the stratum represents the population.
- Stratified sampling is often used when one or more of the stratums in the population have a low incidence relative to the other stratums.

Common Definition on Population

L	**Number of strata**
i	**index for strata** $i = 1 \ldots L$
N_i	**Population count for strata** i
N	**Population count**
	$= \sum_{i=1}^{L} N_i$
μ_i	**Population mean for strata** i
μ	**Population mean**
$\bar{y}_{st}$	**Estimator of** μ

Common Definition on Sampling

τ_i	**Population total for strata** i
τ_i	$= N_i \mu_i$
τ	**Population total**
	$= \sum_{i=1}^{L} \tau_i$
n_i	**Sample count for strata** i
$\bar{y}_i$	**Sample mean for strata** i
$\bar{s}_i^2$	**Sample variance for strata** i

Estimated Mean

Estimator of the population mean μ:

$$\bar{y}_{st} = \frac{1}{N}\sum_{i=1}^{L} N_i \bar{y}_i.$$

Estimated Variance

Estimated variance of $\bar{y}_{st}$:

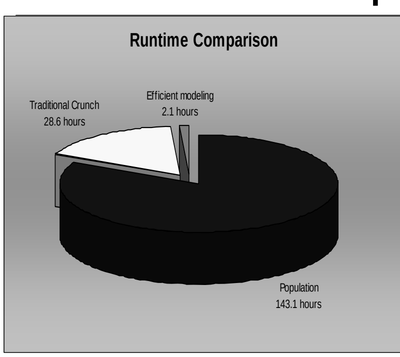
$$\begin{aligned} \hat{V}(\bar{y}_{st}) &= \frac{1}{N^2}\sum_{i=1}^{L} N_i^2 \hat{V}(\bar{y}_i) \\ &= \frac{1}{N^2}\sum_{i=1}^{L} N_i^2 \left(\frac{N_i - n_i}{N_i}\right)\left(\frac{s_i^2}{n_i}\right) \end{aligned}$$

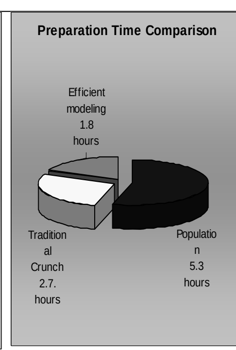
Estimated Error

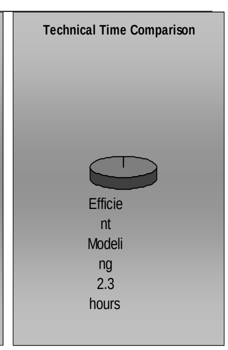
Bound on the error of estimation:

$$2\sqrt{\hat{V}(\bar{y}_{st})} = 2\sqrt{\frac{1}{N^2}\sum_{i=1}^{L} N_i^2 \left(\frac{N_i - n_i}{N_i}\right)\left(\frac{s_i^2}{n_i}\right)}$$

Runtime Comparison

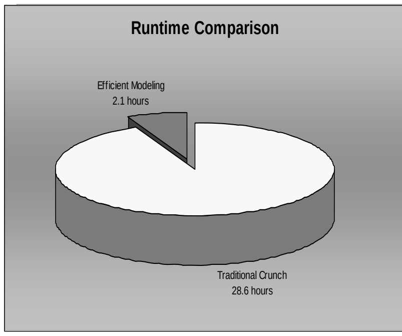


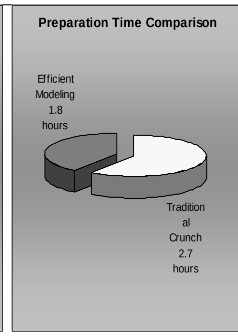


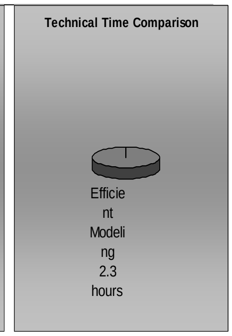


	Population	Crunch	Efficient
Runtime	143.1	28.6	2.1
Preparation	5.3	2.7	1.8
Technical Skill			2.3
Total Time	148.4	31.3	6.2

Runtime Comparison
Traditional Crunching v.s Efficient Modeling

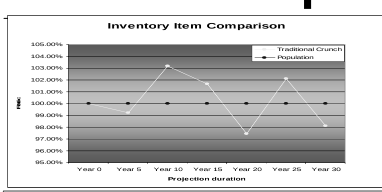


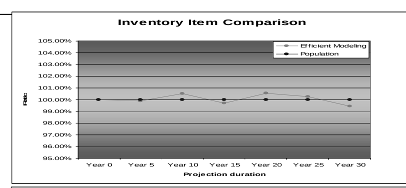


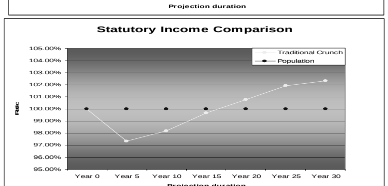


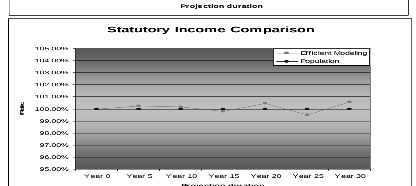
	Population	Crunch	Efficient
Runtime	143.1	28.6	2.1
Preparation	5.3	2.7	1.8
Technical Skill			2.3
Total Time	148.4	31.3	6.2

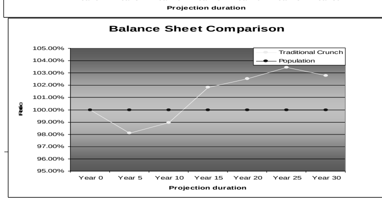
Result Comparison

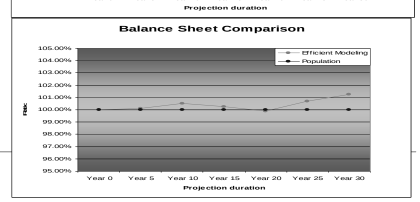












Bibliography

- Kaufman, L. and Rousseeuw, P.J. (1990). Finding Groups in Data: An Introduction to Cluster Analysis.
- Struyf, A., Hubert, M. and Rousseeuw, P.J. (1997). "Integrating Robust Clustering Techniques in S-PLUS," Computational Statistics and Data Analysis, 26, 17-37.

SOA 2008 Spring Meeting
Model Efficiency via Cell Compression
Tuesday, June 17, 2008

Craig A. Roberts, FSA

craig.roberts@milliman.com

Agenda

- Overview of Model Compression Goals
- Compression Techniques
- Real Life Results
- Optimal Compression
- Q&A Session

Compression Goals

- Reduction of Projection Run-Time
  - RBC and PBA requirements are demanding more resources
  - Grid computing is only part of the solution
  - Nested stochastic approaches for pricing increases resource crunch

Compression Goals

- Result Accuracy
  - More compression can mean less accuracy
  - Seriatim approach for all scenarios is impractical
  - Scenarios driving results may change over time

Compression Goals

- Measures of compression success
  - Materially matches Seriatim results
  - Significantly smaller model over prior methods
  - Easily reproducible for future valuation dates
  - Scalable to meet specific compression needs

Basic Compression Approaches

Well understood and common

1. Mid-Year or Mid-Quarter Issue Points
2. Modeled Issue Ages
3. Sex Aggregation with adjusted Female Issue Ages
4. Minor to Major Plan Mapping
5. Fund Modeling
   - Very important when modeling one record/fund
   - Still important when modeling one record/policy

Advanced Compression Approaches

Issue Year Floors

- Only works well for some product types
- Adjust issue age to keep same attained age
- Application of floor limited by:
  - Policy duration assumptions (ex. Surrender charges)
  - Policy issue year assumptions (ex. Valuation Rates)
- A PBA working reserve may remove some limits

Issue Year Floor Example

Issue Year Floor at 1985

Original		Floored	
Issue Year	Issue Age	Issue Year	Issue Age
1980	35	1985	40
1981	35	1985	39
1982	35	1985	38
1983	35	1985	37
1984	35	1985	36
1985	35	1985	35
1986	35	1986	35
1987	35	1987	35

Performed prior to Age Modeling

Advanced Compression Approaches

Dynamic Mortality Modeling

- Potentially 100% accurate aggregate sex modeling
- Female/Male proportion recorded in inforce
  - Proportion by Face (Traditional business), or Account Value
- Proportion updated dynamically by formula
- Mortality, premium rates, etc. applied via proportion
  - May fail to be 100% match for traditional based reserving
  - Working reserve approach bypasses this limitation

Dynamic Mortality Example

Original Cells

Sex	AVIF	Proportion
Male	10000	66.7%
Female	5000	33.3%
Total	15000	100%

New Cell

Sex	AVIF	Female %
Male	15000	33.3%

Dynamic Re-determination of Female Proportion

Policy Yr Qx	Male	Female	Beg. Yr Fem. %	Aggregate Qx	End Yr Fem. %
12	0.03	0.02	33.3%	0.0267	33.6%
13	0.04	0.03	33.6%	0.0366	33.8%
14	0.05	0.04	33.8%	0.0466	34.0%

Advanced Compression Approaches

In-the-Moneyness (‘ITM’) Banding

- Specifically for VA/VUL Guaranteed Benefits
- Allows compression into model points and prevents the in-the-money policies from combining with out-of-the-money
- For multiple benefit types (GMDB vs. GMIB, Ratchet vs. Rollup), the ITM ratio is insufficient.
  - An additional key serves as the “benefit driver”

ITM Banding Example

Driver Table

Driver	Key
N/A	_
Rollup	1
Ratchet	2

ITM Banding Table

Ratio	Key
0%	_
50%	A
100%	B
200%	C
201%+	D

Benefit Value						
AVIF	Rollup	Ratchet	ITM Driver	ITM Ratio	Driver Key	Band Key
5,000	-	-	N/A	0	_	_
5,000	4,000	2,000	Rollup	80%	1	B
10,000	12,000	15,000	Ratchet	150%	2	C
100	1,000	50	Rollup	1000%	1	D

Advanced Compression Approaches

Immaterial Record Reduction

- The premise is that a $1 record takes the same amount of time to process as a $1 Million record
- How can we intelligently determine what is a ‘small record’ vs. one which is material?
- How should these chosen immaterial records be recombined into an aggregate record?

Advanced Compression Approaches

Immaterial Record Reduction: Using Average Ages

1. Generate the expected unadjusted inforce
2. Rank all records in descending order by key result driver(s)
3. Select out records consisting of X% of the total driver value. These will remain unmodified.
4. For the remaining small records, collapse by issue age, assigning a weighted average issue age.

    - Best performance when fewer assumptions rely upon age
    - Best completed prior to any issue age modeling
        - But after every other compression technique

Immateriality Reduction Example

Coverage Ratio set to Top 60% of Value

Prior to Issue Age Averaging

	Benefit Value			Benefit Proportion			
Cell #	AVIF	GMDB	GMIB	AVIF %	GMDB %	GMIB %	Coverage
1	6,000	-	30,000	28%	0%	64%	Material
2	5,000	4,000	2,000	23%	24%	4%	Small
3	10,000	12,000	15,000	47%	71%	32%	Material
4	500	1,000	50	2%	6%	0%	Small
Total	21,500	17,000	47,050	100%	100%	100%	

After Application of Average Issue Age

	Benefit Value			Benefit Proportion		
Cell #	AVIF	GMDB	GMIB	AVIF %	GMDB %	GMIB %
1	6,000	-	30,000	28%	0%	64%
2	5,500	5,000	2,050	26%	29%	4%
3	10,000	12,000	15,000	47%	71%	32%
Total	21,500	17,000	47,050	100%	100%	100%

Testing Compression Results

- Clearly critical to determining success
- Test with tail scenarios
- Allow selection of compression level
  - Coverage Ratio
  - Granularity of In-the-Moneyness bands
- Retest periodically

Compression Results
Real Life Model

Summary of Model Results

ITM Band Compression	Materiality Factor	Ave. Result as a % of Seriatim	Compression Model/ Seriatim	Compression Seriatim/ Model
Low	80%	**99.3%**	4.30%	23.25
Max	80%	**100.3%**	2.07%	48.20
Max	65%	**99.2%**	1.83%	54.65
Max	50%	**97.0%**	1.68%	59.40
	Original Model	**97.5%**	16.90%	5.92
	Seriatim	**100.0%**		

Optimal Compression

Key Components

- Automatic
- Minimal Overhead (in terms of both user and computing resources)
- Recognition of Scenario Variation
- Recognition of Model Structure
- Flexible Selection of Result Driver or Drivers
- Scalability of Compression

Optimal Compression

Approach
- Algorithmic
- The present value of result driver or drivers are summarized for each record and across every scenario
- User specified weights are assigned when multiple drivers are selected
  - Example: Establishing a trade off between reserve fit and profit fit
- Records remodeled into each other based on least squares driver differences
- Policy size can be considered as a separate factor within this algorithm
- Remodeled inforce automatically created
- Development of such a process is currently underway

Discussion

Questions?