# Statistical Sampling Techniques in Actuarial Modeling

**A presentation to the Actuaries' Club of Hartford & Springfield**
**By:**
**Jay Vadiveloo (Senior Consultant, Towers Watson) and**
**Milanthi Sarukkali (Graduate Student, University of Connecticut)**

**May 23, 2011**



TOWERS WATSON

## Presentation Outline

- Background
- Description of Replicated Stratified Sampling (RSS) technique
- Simple illustration to motivate why RSS works
- Single-pay whole life insurance illustration
- Preliminary research findings
- VACARVM pilot test results
- Open research topics

## Background

- Critical need to speed up processing time for complex actuarial calculations
- Need to balance speed of processing with accuracy and stability of results
- Available techniques currently being used include:
  - Grouping techniques
  - Scenario reduction techniques
  - Replicating Portfolio techniques
- Use of statistical sampling techniques, while common in every other discipline, is not used in actuarial modeling work
  - Easy access to the complete inforce of insurance contracts
  - Availability of high speed computers and sophisticated processing techniques
  - Inability to control or estimate sampling error

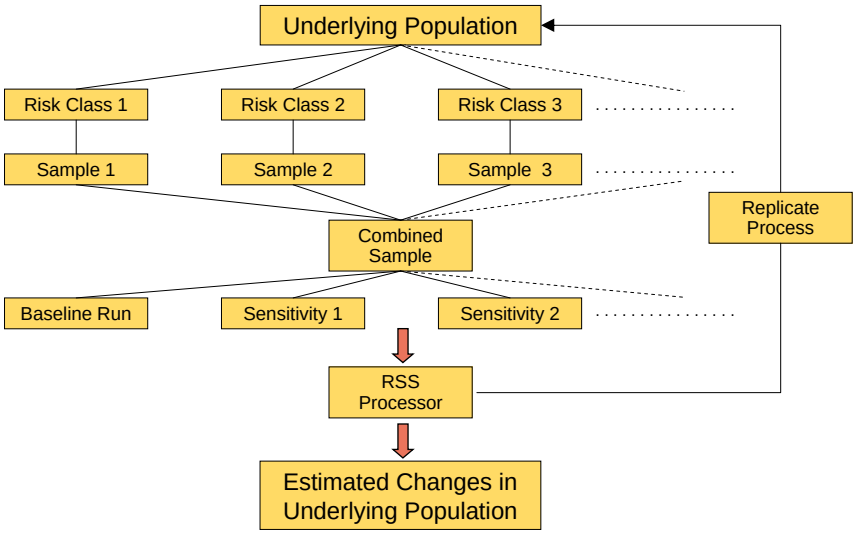




## Replicated Stratified Sampling (RSS)

- Uses stratified sampling to randomly select individual contracts *without replacement* from the inforce
- Replicates (repeats) the process by drawing new random samples from the inforce after replacing all prior samples drawn.
- By replicating the sampling process and then combining samples “cleverly”, sampling error can be measured and controlled
- Processing time to estimate changes in a risk metric because of any set of assumption changes can be reduced *exponentially* using the RSS technique

## RSS Technique







## RSS Patent Application

- Towers Watson is pursuing patent protection for the RSS algorithm
  - Patent application filed in September 2009 with five independent claims, and 30 claims altogether
- One proprietary embodiment of RSS includes drawing repeated samples from a population of financial data and combining the sample information to generate a population estimate
- Additional proprietary features of RSS include:
  - Estimating both the distribution of a population risk metric using the RSS process as well as the change in the distribution of a population risk metric
  - Financial data being modeled using the RSS process, including Life, P&C, hedging, economic, and/or asset modeling applications.
  - Generating samples using stratified sampling versus simple random sampling, elimination of outlier samples, different methods on combining samples, etc.

## Simple Illustration

- Generate 50,000 baseline values from a normal distribution with mean 10 and standard deviation 3. Denote values by X
- For each X value, generate a corresponding shocked value from a normal distribution with mean 5X + 10 and standard deviation 9. Denote these 50,000 shocked values by Y
- We wish to estimate the ratio R = sum(Y)/sum(X)

| Population Statistics | |
|---|---|
| Sum of $X_i$ | 500,782 |
| Sum of $Y_i$ | 3,004,141 |
| **Population Ratio** | **6.00** |





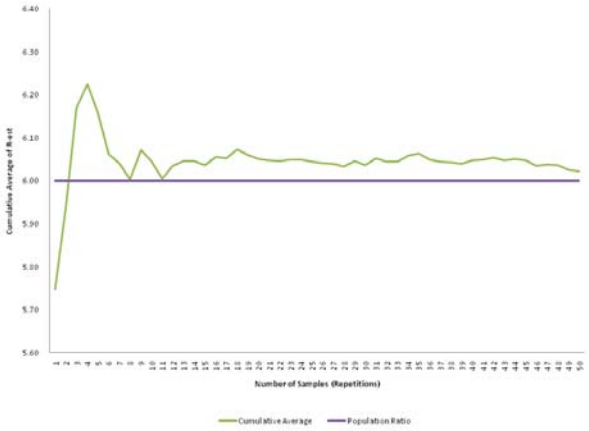
## Simple Illustration – continued

- Randomly select 10 of the (X,Y) pairs from the 50,000 pairs in the population
- Generate the sample ratio R-est as the initial estimate of R

$$R-est = \frac{\sum_{i=1}^{10} y_i}{\sum_{i=1}^{10} x_i}$$

- Replicate the process and calculate the cumulative average of the sample estimates

## Simple Illustration – Results










## Simple Illustration – Conclusions

- When the number of replications equal 50, the average of the sample ratios is within 0.3% of the population ratio
- Total sample of 500 pairs is only 1% of the inforce of 50,000 pairs

## Single pay whole life insurance illustration

- Randomly generate 10,000 insurance policies varying by issue age, face amount and gender
- In the baseline scenario, for each policy, calculate Ax using an industry mortality table and a 5% discount rate
- For the shocked scenario, increase the annual mortality rate by 50% for each policy
- Generate 5,000 times of death for each of the 10,000 policies under both the baseline and shocked scenario
- Generate the population distribution of the sum of Ax's for the 10,000 policies under both the baseline and shocked scenarios.
- We are interested in estimating R = ratio of 90-th percentiles of the shocked and baseline populations

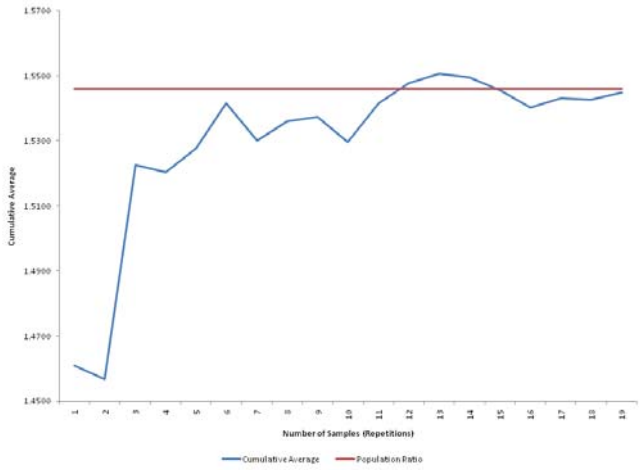




## Single pay whole life insurance illustration – continued

- Randomly sample 10 policies from the 10,000 inforce policies
- Use the *same* 5,000 times of death for the 10 policies that were generated for the population under both the baseline and the shocked scenario
- Generate the sample distribution of the sum of the Ax's for the 10 policies under both the baseline and shocked scenarios
- Generate R-est = ratio of the 90-th percentiles of the shocked and baseline sample distributions
- Replicate the process and calculate the cumulative average of R-est

## Single pay whole life insurance illustration – results








## Single pay whole life illustration – Conclusions

- When the number of replications equal 19, the average of the sample ratios is 0.1 % of the population ratio
- Total sample of 190 pairs is less than 2% of the inforce of 10,000 policies

## Preliminary research findings

- While using simple random sampling with the RSS technique does work, stratified sampling using broad risk classes improves the speed of convergence of RSS
- The risk classes have less to do with standard grouping techniques, but are more related to homogeneity in the ratio of the change in the risk metric being estimated.
- The speed of convergence of the RSS technique is *independent* of the size of the underlying inforce population.
- *More* replications with a *smaller* sample produces *faster* convergence compared to *less* replications with a *larger* sample. This is only true after a *minimum* sample size is achieved.
- RSS convergence can be significantly improved if *outlier* sample ratio estimates are systematically removed from the set of replicated sample ratios





## RSS Pilot Study

- RSS technique applied to a variable annuity block of a major life insurance company.
- Analyzed impact of
  - an ***immediate*** 15% drop in equity funds on VACARVM reserves
  - an ***immediate*** 35% drop in equity funds on VACARVM reserves
- Analysis done for 3 legal entities both ***before and after reinsurance***
- Analysis compared the ***change*** in the VACARVM reserve in the population versus using the RSS technique on 50, 100, 150 and 200 samples of 30 policies each
- Error rate defined as: $\frac{|A-B|}{B}$

where A = change in VACARVM reserve using the RSS technique
B = change in VACARVM reserve in the population

# RSS Results – Sensitivity 1

After Reinsurance

| Legal Entity | # Of samples | RSS Ratio | POP Ratio | Error Rate |
|---|---|---|---|---|
| 1 | *50* | 1.513 | 1.665 | 9.14% |
| | *100* | 1.615 | | 3.00% |
| | *150* | 1.640 | | 1.52% |
| | *200* | 1.667 | | 0.12% |
| 2 | *50* | 1.611 | 1.442 | 11.73% |
| | *100* | 1.486 | | 3.10% |
| | *150* | 1.424 | | 1.25% |
| | *200* | 1.450 | | 0.59% |
| 3 | *50* | 3.015 | 3.634 | 17.02% |
| | *100* | 3.482 | | 4.18% |
| | *150* | 3.735 | | 2.78% |
| | *200* | 3.633 | | 0.02% |

Before Reinsurance

| Legal Entity | # Of samples | RSS Ratio | POP Ratio | Error Rate |
|---|---|---|---|---|
| 1 | *50* | 2.067 | 2.113 | 2.17% |
| | *100* | 2.124 | | 0.51% |
| | *150* | 2.106 | | 0.32% |
| | *200* | 2.110 | | 0.15% |
| 2 | *50* | 1.826 | 1.821 | 0.25% |
| | *100* | 1.812 | | 0.49% |
| | *150* | 1.818 | | 0.17% |
| | *200* | 1.822 | | 0.03% |
| 3 | *50* | 2.853 | 3.179 | 10.24% |
| | *100* | 3.028 | | 4.74% |
| | *150* | 3.153 | | 0.82% |
| | *200* | 3.177 | | 0.07% |





# RSS Results – Sensitivity 2

After Reinsurance

| Legal Entity | # Of samples | RSS Ratio | POP Ratio | Error Rate |
|---|---|---|---|---|
| 1 | *50* | 2.771 | 4.236 | 34.58% |
| | *100* | 3.783 | | 10.68% |
| | *150* | 4.071 | | 3.88% |
| | *200* | 4.216 | | 0.45% |
| 2 | *50* | 1.117 | 1.505 | 25.75% |
| | *100* | 1.339 | | 11.00% |
| | *150* | 1.386 | | 7.89% |
| | *200* | 1.506 | | 0.10% |
| 3 | *50* | 17.574 | 25.148 | 30.12% |
| | *100* | 23.239 | | 7.59% |
| | *150* | 26.601 | | 5.78% |
| | *200* | 25.225 | | 0.31% |

Before Reinsurance

| Legal Entity | # Of samples | RSS Ratio | POP Ratio | Error Rate |
|---|---|---|---|---|
| 1 | *50* | 5.575 | 7.322 | 23.86% |
| | *100* | 7.043 | | 3.81% |
| | *150* | 7.099 | | 3.04% |
| | *200* | 7.330 | | 0.10% |
| 2 | *50* | 4.884 | 4.275 | 14.24% |
| | *100* | 5.117 | | 19.70% |
| | *150* | 4.652 | | 8.81% |
| | *200* | 4.282 | | 0.15% |
| 3 | *50* | 13.845 | 20.354 | 31.98% |
| | *100* | 19.059 | | 6.37% |
| | *150* | 21.203 | | 4.17% |
| | *200* | 20.336 | | 0.09% |

## Modeling Efficiency of RSS Technique

- Using 200 samples of either 20 or 30 policies per sample, the RSS technique captured changes in AG43 reserves to within 1% of the "true" population change

  - Depending on the number of policies in each legal entity, the potential reduction in processing time ranged from 98.9% to 99.6%





## Open research topics

- Developing theoretical proofs for the convergence of the RSS technique and publishing the results in an academic actuarial or statistical journal
- Analyzing the optimal number of risk classes to create to maximize the speed of convergence of RSS
- Developing an optimal stopping time technique on the number of replications to generate without knowing the value of the population ratio being estimated
- Expanding the RSS technique to estimate a population risk metric and not just the change in a population risk metric
- Use of the RSS technique in hedging calculation to generate better hedge estimates by increasing the number of stochastic simulations for each sample.
- Expanding the RSS technique to P&C applications and other financial calculations