LFV-133: Cluster Analysis: A Spatial Approach to Actuarial Modeling

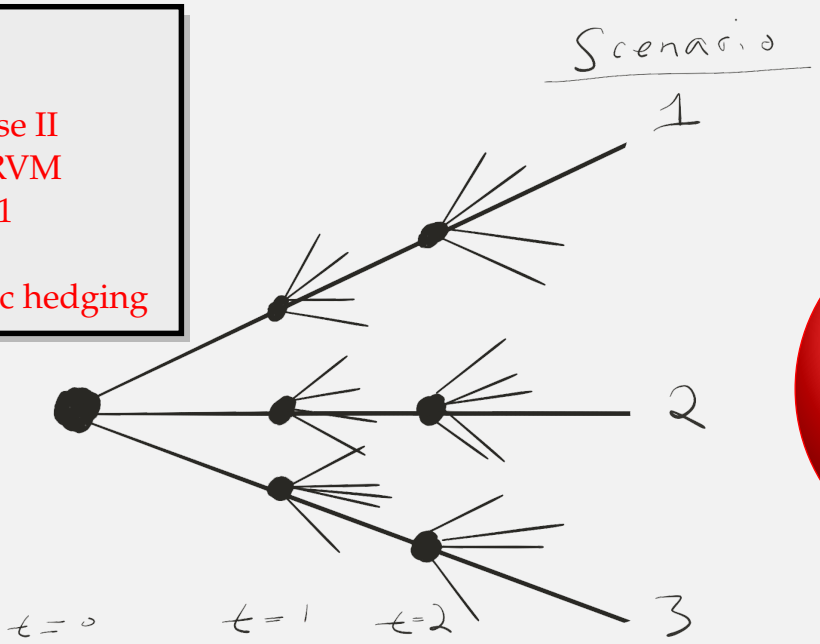
Freedman and Reynolds (August 2008)

Video By: J. Eddie Smith, IV, FSA, MAAA

Key Exam Topics in This Lesson

The Problem: Nested Stochastic Modeling

- IFRS
- PBA
- C-3 Phase II
- VA CARVM
- SOP 03-1
- FAS 133
- Dynamic hedging



Dramatically increased runtime

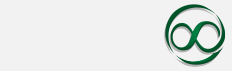
Classic Approaches to Reducing Runtime

1. Give us **FASTER computers!**
2. Now give us **MORE computers!**
3. Let's run **fewer scenarios**
   - Or **fewer paths**
   - Or **fewer shocks**
4. Let's do **less frequent re-balancing**

} Less accuracy too

Or let's model fewer cells…

Classic Modeling: Modified Seriatim



Seriatim ⇒ no "modeling" done ⇒ include all policies as-is

Modified seriatim methods attempt to reduce the volume of data modeled

1. Combine policies with same issue month, plan, premium mode, etc.
2. Use quinquennial or decennial issue ages

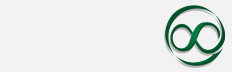
} 10:1

3. **Combine risk classes or map minor plans into major plans**

   Commonly done, but has **drawbacks**:

   3.1 Must know something about minor plans
   3.2 Mapping rules are subjective, hard to automate
   3.3 Must update rules for new plans and as in-force changes
   3.4 Projected values may not be valid
   3.5 Hard to apply rules for multiple life policies and investment guarantees

LFV-133: Cluster Analysis – A Spatial Approach



Cluster Modeling: User Steps

1. **Define location variables and their weights**
   - Reserves, CSV, premium, PV guaranteed benefits, PV profit, $\sum$ Premiums
   - Higher weight $\Rightarrow$ higher priority
2. **Define a size variable** (face, AV)
   - Smaller policies get mapped first
3. **Define segments** that should not be mapped across
   - Business that should not be mixed: plan code, issue year or GAAP era, etc.
   - Reduces runtime
     - Runtime for 10 equal segments = $\frac{1}{10} \times$ runtime for whole group
   - Other reasons: reporting, reconciliation, LVs don't sufficiently distinguish policies
4. **Specify a target number of clusters**

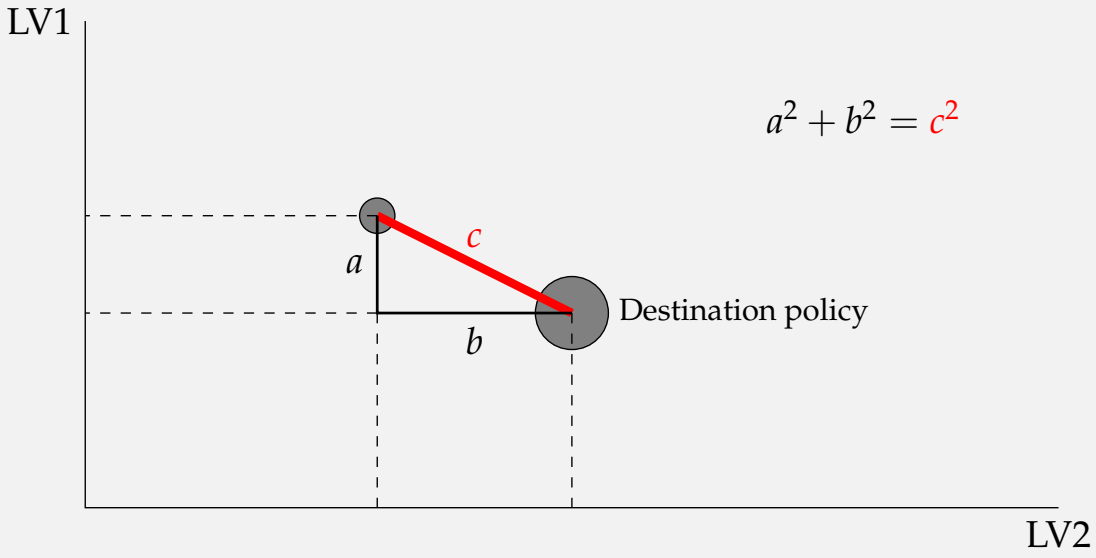
Cluster Modeling: Process Steps

1. **Calculate distance between any 2 policies**
   - n-dimensional sum-of-squares approach

     3 LV case: $\sqrt{(LV1_1 - LV1_2)^2 + (LV2_1 - LV2_2)^2 + (LV3_1 - LV3_2)^2}$

   - Normalize LVs by dividing by their size-weighted standard deviation
2. **Determine importance of each policy**
   - Importance = policy size $\times$ distance from the nearest policy
   - Higher importance $\Rightarrow$ less likely to be mapped
3. **Create clusters**
   - Map policy with lowest importance to its nearest neighbor
   - Process continues until cluster target is met
4. **Determine representative policy for each cluster**
   - A single policy that is closest to cluster's average location
   - Gross up to reflect size of cluster

Visualizing Euclidean Distance in 2 Dimensions



Cluster Model Advantages Over Classic

1. Applies to liabilities or even assets
2. Far better compression ratios for a given model-to-actual fit
3. Easily automated $\Rightarrow$ less manual effort
4. Can be maintained and applied at later valuation dates
5. Priority measures of model fit measures can be customized
6. Applies to seriatim in-force or to modeled in-force
7. Model points are easily adjusted to change for desired granularity
8. Allows on-the-fly analysis of model fit without rerunning a model

Potential Applications of Cluster Modeling

1. **Medium-sized models**
   - Replacement for classic models: nearly reproduces seriatim results
   - Uses a large number of segments or cells per segment
   - Uses typical LVs: issue age, issue year, in-the-moneyness, etc.
2. **Small models**
   - Can also reproduce seriatim results but with less accuracy
   - Good for estimating CTEs
   - May not be accurate enough for tail analysis
3. **Very small models**
   - Not appropriate for tail analysis
   - May be used to process many scenarios to select a smaller set for another model
   - Could be used to quantify sensitivities to key assumptions

As model size falls, use more care in validating!

Cluster Modeling Applied to Traditional and Term Life

Traditional Life/Health Model

- LVs: reserve, FY premiums, FY claims, PV profits
- Segments: None
- Compression: 120,000 cells to 200 cells (60:1)
- Higher-weighted LVs nearly identical
- Close match (95–96%) on lower weighted LVs

Term Life Model

- LVs: reserves, PV various cash flow buckets, $\sum$ Premiums for various buckets
- Segments: Issue year-based, and LT period
- Compression: 1.1MM policies to 10,000 cells and 300 cells (110:1 and 3667:1)
- Results: both cluster models fit very well

Cluster Modeling Applied to a VA Block

"A VA model is only useful if it performs well under a variety of scenarios"

VAs are difficult to compress using traditional mapping

- Similar policies have different in-the-moneyness
- Policyholder behavior drives investment allocation $\Rightarrow$ future returns

Authors tested 2 cluster models against a 9000-cell classic model

1. 250-cell model
   - Fit very well
   - Could be used for tail analysis
2. 50-cell model
   - Also fit well, but had more deviations
   - Not appropriate for tail analysis

LVs: AV, PV DBs and other benefits, PV profits (highest weight)

Validating a Small(er) Model

1. **Run a large model to obtain results for comparison**
2. **Do a static validation**
   - Compare starting balance sheet values
   - Adjust weights as needed, make tradeoffs
3. **Do a dynamic validation**
   - Compare projected income statement values
   - Compare all components if replacing the large model
4. **Run small model over a small scenario set**
   - Select scenarios of importance
   - May only be needed initially and occasionally in the future