Session 2

Predictive Analytics in Policyholder Behavior

Eileen Burns, FSA, MAAA
David Wang, FSA, FIA, MAAA
Predictive Analytics in Policyholder Behavior

Eileen Burns, FSA, MAAA
Principal & Consulting Actuary
Milliman Inc.

David Wang, FIA, FSA, MAAA
Principal & Consulting Actuary
Milliman Inc.

27th August 2018
Agenda

- Current state in life and annuity
- Examples of where predictive analytics helps
- Implication on assumption setting process
- Interesting applications

Eileen Burns, FSA, MAAA
Principal & Consulting Actuary
Seattle
Eileen.Burns@milliman.com

Education and Qualifications
Masters

Lawrence University (1998 - 2002)
BA, Mathematics

Current responsibilities
- Principal on Milliman’s data analytics team
- Product manager for Recon, a Milliman predictive analytics and data product targeted at enhancing experience analysis
- Vice-chair of SOA Predictive Analytics and Futurism section

David Wang, FIA, FSA, MAAA
Principal & Consulting Actuary
Seattle
David.Wang@milliman.com

Education and Qualifications
University of California at Berkeley, HAAS School of Business (2005 - 2006)
MFE, Financial Engineering

Nanyang Technological University (1994 - 1998)
B. Business

Current responsibilities
- Co-leads Milliman’s team specializing in applying data analytics to assist the life and annuity industry in the United States.
- Co-leads Milliman life consulting practice in Seattle
Current State in Life and Annuity
What is Predictive Analytics and Predictive Modeling

**Predictive analytics** uses many techniques from data mining, statistics, modeling, machine learning, and artificial intelligence to analyze current data to make predictions about future. **Predictive modeling** is a process used in predictive analytics to create a statistical **model** of future behavior. (Google Search)
Policyholder Behavior Modeling: Progression of States

- **Traditional State**
  - Traditional one-way actuarial techniques to estimate behavior by age/duration and limited number of other characteristics using experience where it exists
  - Primarily macro-oriented... little use of detailed information on policyholder characteristics
  - Judgment and guesswork where experience does not exist
  - Next-generation experience studies using policyholder longitudinal data.
  - Use much wider set of explanatory variables readily available to company
    - Internal data (Product features, distribution channel, policyholder and contract characteristics)
    - Macro data (Economic data, financial market conditions)
  - More sophisticated analysis techniques to find non-linear, multivariate effects, complex interactions
  - Employ external consumer/financial/health and big/unstructured data sources in a full Predictive Analytics framework.
  - Develop individual policyholder profiles
Applications of Predictive Analytics in Life and Annuity

<table>
<thead>
<tr>
<th>Behavior modeling</th>
<th>Beyond</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Lapse (mostly annuity)</td>
<td>• Predictive underwriting</td>
</tr>
<tr>
<td>• Withdrawal (variable annuity)</td>
<td>• Target market and lead generation</td>
</tr>
<tr>
<td>• Post level term shock lapse</td>
<td>• Producer recruitment and retention</td>
</tr>
<tr>
<td>• Premium persistency</td>
<td>• Cross-sell</td>
</tr>
<tr>
<td>• Fund transfer</td>
<td></td>
</tr>
</tbody>
</table>

Actuarial  Data Analytics
Examples of Where Predictive Analytics Helps
Improve Predictions

Overall Improvement in Predictions

Relative impact from predictors
Test Hypothesis and Answer Question

• Is there a difference in sensitivity to crediting spread among distribution channels?

• Does the MVA effectively eliminate sensitivity to crediting spread?

**FA: Spread sensitivities by channel**

- Channel
  - Bank
  - Broker dealer
  - Captive agent
  - Independent agent

**MVA: Spread sensitivities by channel**

- Channel
  - Broker dealer
  - Captive agent
  - Independent agent
Identify drivers

- Previous behavior – e.g. withdrawal behavior
- People – demographics and distribution channel
- Product design – MVA, surrender charge structure, guaranteed minimum
- Macroeconomics – market rates, unemployment
Confidence Intervals

Model predictions and confidence bands versus actual experience

Baseline model

Full model

Quarterly lapse rates

Confidence Intervals

Actual lapse rate
Predicted lapse rate
95% Confidence interval
Implication on Assumption Setting Process
Typical predictive modeling process

Data Preparation

Data Analysis

Model Building

multiple files

PM data

Millions of records

Hundreds of variables

training/holdout test
Era of Big Data has come, but Life Insurers Need to Catch Up!

Challenges the life insurance industry faces:

- Little systematic collection and storage of data
- Legacy system inadequate for new data analytics
- Limited data to differentiate customer
- Silos still exist
Data visualization is more than just better pictures

- More data, more information, more dimensions, calls for better visualization
- Makes traditional date reporting inefficient
- Provides guidance and tips on how predictive models should be built
Bring predictive model in assumption setting process

**Implementation**
Can we model all the predictive drivers in the actuarial cash flow projection?
If not, how do we make compromise and recognize the loss of accuracy.

**Validation**
How is the goodness of fit over different dimensions?
How are we comfortable with confidence intervals?
Domain knowledge is essential to make sense of results.

**Communication**
How do actuaries convince themselves and management that PM is needed?
How do actuaries communicate model results to senior management?

**Control & Governance**
Predictive modeling requires new controls & governance. How do we develop appropriate standards?
Who is qualified to review and sign off?
What type of documentation should be retained?
Some Interesting Applications
# Evaluation of behavioral tail risk

## Types of lapse tail risk

<table>
<thead>
<tr>
<th>Drift</th>
<th>Diffusion</th>
<th>Extreme Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk that best estimate lapse rates vary under different market conditions</td>
<td>Risk that estimates of the entire lapse function are off</td>
<td>Risk that some unprecedented events may impact lapse in an extreme way</td>
</tr>
<tr>
<td>Captured by a dynamic lapse component</td>
<td>Captured by simulation of lapse behaviour using predictive model</td>
<td>Resort to some manner of judgement call</td>
</tr>
</tbody>
</table>

---

**Drift**
- Risk that best estimate lapse rates vary under different market conditions
- Captured by a dynamic lapse component

**Diffusion**
- Risk that estimates of the entire lapse function are off
- Captured by simulation of lapse behaviour using predictive model

**Extreme Event**
- Risk that some unprecedented events may impact lapse in an extreme way
- Resort to some manner of judgement call
Lapse behavior simulation

\[ \text{logodds} \sim a + b_1 \times \text{Variable1} + b_2 \times \text{Variable2} + b_3 \times \text{Variable3} + \varepsilon \]

Model Assumptions:

- Each coefficient $b_x$ is normally distributed
- The error term $\varepsilon$ is normally distributed with a mean of 0
- Correlation between each predictive $\text{VariableX}$ can be given by a correlation matrix
- The standard deviation of $\varepsilon$ denoted by $\Omega$ can be determined from the correlation matrix using numerical methods such as Cholesky decomposition
Lapse behavior simulation – Determine best estimate

• Consider the following model, where the only predictive variable considered is In-The-Moneyness

\[
\text{logodds} \sim a + b_1 \times \text{ITM} + \varepsilon
\]

• After fitting your experience to the model, the following best estimate calibration is attained:

\[
\text{logodds} = 0.5 + (-2) \times \text{ITM} + 0
\]

Resulting best estimate lapse rate (\(p\)):

<table>
<thead>
<tr>
<th>ITM</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>225%</td>
<td>1.8%</td>
</tr>
<tr>
<td>175%</td>
<td>4.7%</td>
</tr>
<tr>
<td>125%</td>
<td>11.9%</td>
</tr>
<tr>
<td>75%</td>
<td>26.9%</td>
</tr>
<tr>
<td>25%</td>
<td>50.0%</td>
</tr>
</tbody>
</table>
Lapse behavior simulation – Simulating the risk of model misestimation

- Alternatively, we can simulate lapse rates by allowing the coefficients to vary according to their standard deviation, assuming a multivariate normal distribution

\[ \text{logodds} = 0.5 + N_1(0, \Omega) + (-2 + N_2(0, \Omega)) \times \text{ITM} \]

<table>
<thead>
<tr>
<th>( \varepsilon(i) )</th>
<th>ITM</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>225%</td>
<td>1.8%</td>
</tr>
<tr>
<td>0</td>
<td>175%</td>
<td>4.7%</td>
</tr>
<tr>
<td>0</td>
<td>125%</td>
<td>11.9%</td>
</tr>
<tr>
<td>0</td>
<td>75%</td>
<td>26.9%</td>
</tr>
<tr>
<td>0</td>
<td>25%</td>
<td>50.0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \varepsilon = {-0.2, -0.1} )</th>
<th>ITM</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \varepsilon )</td>
<td>225%</td>
<td>1.2%</td>
</tr>
<tr>
<td>-0.2, -0.1</td>
<td>175%</td>
<td>3.3%</td>
</tr>
<tr>
<td>-0.2, -0.1</td>
<td>125%</td>
<td>8.9%</td>
</tr>
<tr>
<td>-0.2, -0.1</td>
<td>75%</td>
<td>21.8%</td>
</tr>
<tr>
<td>-0.2, -0.1</td>
<td>25%</td>
<td>44.4%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( \varepsilon = {0.2, 0.1} )</th>
<th>ITM</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \varepsilon )</td>
<td>225%</td>
<td>2.8%</td>
</tr>
<tr>
<td>0.2, 0.1</td>
<td>175%</td>
<td>6.8%</td>
</tr>
<tr>
<td>0.2, 0.1</td>
<td>125%</td>
<td>15.8%</td>
</tr>
<tr>
<td>0.2, 0.1</td>
<td>75%</td>
<td>32.6%</td>
</tr>
<tr>
<td>0.2, 0.1</td>
<td>25%</td>
<td>55.6%</td>
</tr>
</tbody>
</table>
Data-driven segments identify policyholders likely to behave in similar ways

- Number and defining characteristics of segments will be specific to the particular dataset
- Likely defining values for segments include credit score, income, home value, home mortgage loan-to-value, etc.

Segment specific behavior modeling reveals how people use insurance differently

Unsegmented
Differentiation between policyholder behavior and corresponding profitability

- Plots show profitability differences driven purely by behavioral difference due to belonging to different segments.

- Help identify groups of people whose needs are not served properly by current product offerings and identify need for new products
A new perspective of product profitability

Show profitability at state level, or at county or zip code view
Final thoughts
Goal and elements of predictive analytics in policyholder behavior

To predict (individual) policyholder behavior by applying rigorous statistical techniques to large amounts of data under the guided framework designed by subject experts.