PREDICTIVE MODELS IN LIFE INSURANCE

Date: 17 June 2010

Philip L. Adams, ASA, MAAA

Agenda

1. Predictive Models Defined
2. Predictive models past and present
3. Actuarial perspective
4. Application to Life Insurance: ILEC 02-04 and 05-07 data
What is predictive modeling?

“Predictive modeling is the use of modern statistical methods to identify and measure past associations for the sake of inferring associations about the past, the present and the future. Usually, this is expressed through probabilities, and it has a very strong Bayesian flavor.”

-Your humble speaker

This sounds a lot like what actuaries have been doing for a century (except for the “modern statistical methods” part).
What it’s not

- Panacea
- Best thing since sliced bread
- The tool to end all tools

Nuts and Bolts of Predictive Modeling

- Data mining: Use statistical methods to find patterns
  - “Correlation does not imply causation”
    - …but we often have to assert causation
  - Some associations cannot be measured using correlation
    - Correlation is measure of statistical collinearity
    - Correlation implies dependency, but not conversely
    - Example: If $X$ is uniform on (-1,1), then $X^2$ is uncorrelated with $X$
- Take action with those patterns
  - Build a better product or service
  - Incorporate into risk management
PREDICTIVE MODELS

PAST AND PRESENT

PREDICTIVE MODELS: The Past
Horoscopes
PREDICTIVE MODELS: In Fiction
The Witch in Monty Python’s Holy Grail

- “Experience” taught knight that witches float, so compare candidate to other things that float
- Peasants’ suggested comparisons: wood, very small rocks, churches, water fowl
- Analogous to a decision tree

PREDICTIVE MODELS: Real Data, Bad Analysis
Pirates and Global Warming

- Number of pirates has decreased since 1860
- Globe has consistently warmed
- Ergo, Somalia is at the forefront of the fight against global warming
Title of presentation and name of speaker

PREDICTIVE MODELS: Dubious conclusions

More pirates!

Title of presentation and name of speaker

PREDICTIVE MODELS: The Present

- Mortality tables
- Lapse tables
- Trends
- Credit scoring
- Underwriting manuals
- UCS
- Fraud detection
### Comparisons

<table>
<thead>
<tr>
<th>Life Actuarial Modeling</th>
<th>Predictive Modeling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual, ad hoc</td>
<td>Formal, statistically sound systems</td>
</tr>
<tr>
<td>Inefficient use of data</td>
<td>Extracts a lot of information</td>
</tr>
<tr>
<td>Inefficient models</td>
<td>Models efficiently capture relationships</td>
</tr>
<tr>
<td>Modeler biases strong (often hidden)</td>
<td>Modeler biases must be explicit</td>
</tr>
<tr>
<td>…but gets the job done</td>
<td>…remains mysterious</td>
</tr>
</tbody>
</table>

### Analogues

<table>
<thead>
<tr>
<th>Life Actuarial Modeling</th>
<th>Predictive Modeling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whitaker-Henderson Graduation</td>
<td>Additive models with splines</td>
</tr>
<tr>
<td>Analysis of A/E ratios</td>
<td>Regression models with offsets</td>
</tr>
<tr>
<td>Underwriting rules</td>
<td>Decision trees</td>
</tr>
<tr>
<td>Credibility (basic)</td>
<td>Credibility (advanced)</td>
</tr>
</tbody>
</table>
Modeling the ILEC 02-04 Data

- Death claims and exposures on fully underwritten lives from 35 companies collected between 2002 and 2004
- Nearly 700,000 claims overall
- Almost 500,000 in durations 21+
- For this presentation, we confine our attention to a particular subset
  - Attained ages 20-84, durations 1-10
  - Face amounts at least $10,000 and summarized into two bands
  - Left with 37,087 death claims
- We anchor to PPR ultimate model from Edwalds, Craighead, Adams paper
Modeling the ILEC 02-04 Data

Smoker Differential by Attained Age

Band Differential by Attained Age
Modeling the ILEC 02-04 Data

Band Differential by Attained Age - Non-smokers

Band Differential by Duration
Linear Model
• Standard regression model using least squares minimization of some response against some collection of predictor variables
  \[ y = \beta_0 + \beta_1 x + \epsilon \]

Generalized Linear Model
• Expand universe of response variables to probabilities, counts, etc.
  \[ g(y) = \beta_0 + \beta_1 x + \epsilon \]

Generalized Additive Model
• Expand universe of predictors to include smooth functions
  \[ g(y) = \beta_0 + f(x) + \epsilon \]
Modeling with GAMs

- Splines underly "smooth functions"
  - Cubic splines
  - Penalized splines
  - Thin-plate regression splines
  - Tensor product splines

Fitting GAMs to ILEC 02-04:
Non-smoker Surface
Fitting GAMs to ILEC 02-04: Smoker Surface

<table>
<thead>
<tr>
<th>Parameters (offset from ultimate)</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>p-value</th>
<th>e.d.f.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.317</td>
<td>0.062</td>
<td>0.00%</td>
<td>1</td>
</tr>
<tr>
<td>Males</td>
<td>-0.059</td>
<td>0.012</td>
<td>0.00%</td>
<td>1</td>
</tr>
<tr>
<td>Face Amounts $50,000+</td>
<td>-0.399</td>
<td>0.012</td>
<td>0.00%</td>
<td>1</td>
</tr>
<tr>
<td>Non-smokers</td>
<td>-0.298</td>
<td>0.062</td>
<td>0.00%</td>
<td>1</td>
</tr>
<tr>
<td>Smokers</td>
<td>0.488</td>
<td>0.065</td>
<td>0.00%</td>
<td>1</td>
</tr>
<tr>
<td>Non-smokers in Duration 1</td>
<td>-0.823</td>
<td>0.078</td>
<td>0.00%</td>
<td>1</td>
</tr>
<tr>
<td>Non-smokers in Duration 2</td>
<td>-0.199</td>
<td>0.039</td>
<td>0.00%</td>
<td>1</td>
</tr>
<tr>
<td>Smokers in Duration 1</td>
<td>0.000</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Smokers in Duration 2</td>
<td>-0.173</td>
<td>0.055</td>
<td>0.16%</td>
<td>1</td>
</tr>
<tr>
<td>Smooth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-smokers</td>
<td>63 parameters</td>
<td>0.00%</td>
<td>18.06</td>
<td></td>
</tr>
<tr>
<td>Smokers</td>
<td>63 parameters</td>
<td>0.00%</td>
<td>9.81</td>
<td></td>
</tr>
</tbody>
</table>
Fitting GAMs to ILEC 02-04: Smoker Penalty

But can it predict?

- Attach to ILEC 05-07 Dataset
  - Confined to data similar to fitting set
  - Data summarized, so judgment used for choice of age and duration
  - Table projected forward to 2008 using 1% per annum flat
- Compare to 2008 VBT
  - Represents actuarial approach
## 2008 VBT and GAM Compared

### Gender and Smoker Status

<table>
<thead>
<tr>
<th></th>
<th>A/E by Count</th>
<th>A/E by Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Claims</td>
<td>08VBT Primary</td>
</tr>
<tr>
<td>Female</td>
<td>11,402</td>
<td>110.1%</td>
</tr>
<tr>
<td>Non-Smoker</td>
<td>8,812</td>
<td>108.8%</td>
</tr>
<tr>
<td>Smoker</td>
<td>2,590</td>
<td>115.1%</td>
</tr>
<tr>
<td>Male</td>
<td>20,654</td>
<td>107.8%</td>
</tr>
<tr>
<td>Non-Smoker</td>
<td>15,663</td>
<td>106.3%</td>
</tr>
<tr>
<td>Smoker</td>
<td>4,991</td>
<td>113.1%</td>
</tr>
<tr>
<td>Overall</td>
<td>32,056</td>
<td>108.7%</td>
</tr>
</tbody>
</table>

Bottom line results are comparable.

### Face Amount Band

<table>
<thead>
<tr>
<th></th>
<th>A/E by Count</th>
<th>A/E by Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Claims</td>
<td>08VBT Primary</td>
</tr>
<tr>
<td>10,000-24,999</td>
<td>4,062</td>
<td>150.7%</td>
</tr>
<tr>
<td>25,000-49,999</td>
<td>4,212</td>
<td>135.5%</td>
</tr>
<tr>
<td>50,000-99,999</td>
<td>5,531</td>
<td>121.4%</td>
</tr>
<tr>
<td>100,000+</td>
<td>18,251</td>
<td>95.3%</td>
</tr>
<tr>
<td>Overall</td>
<td>32,056</td>
<td>108.7%</td>
</tr>
</tbody>
</table>

GAM has better control for face amount variation.
References


Image Credits

- Wikimedia Commons
- Monty Python
- Veganza.com
- Wenger
- CDC