Predictive Modeling Techniques in Insurance

Tuesday May 5, 2015

JF. Breton – Application Engineer
Opening

- **Presenter:**
  - **JF. Breton:** 13 years of experience in predictive analytics and risk management in insurance and banking

- **This session:**
  - Cover different best-practice predictive modeling techniques in insurance
  - Show how these can answer practical business questions with MATLAB

- **Takeaways:**
  - Big data and new available technology are creating new opportunities in the insurance industry
  - MATLAB can quickly help you get value from your data with its predictive analytics capability
Agenda

- Predictive modeling background
- Case study – Claim settlements forecasting
- Conclusion

(Warning: Crowded and busy slides ahead!)
What is predictive modeling?

- Use of mathematical language to make predictions about the future
- More of an art than a science (lots of trial and error)
Predictive modeling applications in insurance

- Price risks
- Identify target clienteles
- Predict fraud
- Claims triage
- Set assumptions and reserves
- And many more reasons
Where does it fit?
Trends that drive the use of these techniques

- Available technology and large amount of data
- Increased need for customized products/services
- Pressure on top and bottom lines of income statement

Historical perspective: P&C vs. Life & Annuity

- P&C industry has matured much faster than Life & Annuity
- Credit and risk scores have been used to predict future P&C claims for over 20 years
- Short duration P&C products have limited tail risk compared to most life contracts
- Mortality studies can require several years of data to analyze but Life & annuity insurers are catching up
2013 Insurance predictive modeling survey

**Impacts**
- Predictive models now widely used yet still mainly in pricing and underwriting
- Benefits seen on profitability, risk reduction and operational efficiency

**Challenges**
- Lack of sufficient data attributes and skilled modelers
- Data prep and model deployment can often take +3 months each
- Big Data is currently mainly leveraged by large insurers

(Source: Earnix)
The future: Increased focus on these techniques?

- Disruptions may occur from outside the industry
- From risk pooling to “Big Mother”

Disruption From Non-Traditional Insurance Providers

- Application completed
  - Predictive Model Determines UW class
  - Life Insurance Policy Issued
  - Social Media Data
  - Geospatial Data

Predictive Model run annually on all policies

- Health /Lifestyle feedback Provided to p/h
  - Identify lifestyle based risks after policy underwritten and issued
  - Positive feedback/coaching included in contract

Policyholder Chooses to incorporate feedback

- Premium Adjusted / Lapse Decision
  - Reduce tail risk with adjustable premium
  - Policyholder incentive to reduce risk
Some applications and associated techniques

- **Predicting loss severity** *(parametric)*
  - Generalized Linear Model

- **Predicting S&P 500 for annuities** *(time series)*
  - ARIMA modeling
  - GARCH modeling

- **Predicting customer response** *(non-parametric)*
  - Classification learning techniques
  - Measure accuracy and compare models
Our focus on today: Learning Techniques

<table>
<thead>
<tr>
<th>Supervised Learning</th>
<th>Unsupervised Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>(The target is known)</em></td>
<td><em>(The target is unknown)</em></td>
</tr>
<tr>
<td><strong>Parametric (Statistical)</strong></td>
<td></td>
</tr>
<tr>
<td>• Linear Regression</td>
<td>• Cluster Analysis (i.e. K-means)</td>
</tr>
<tr>
<td>• Time Series</td>
<td>• Principal Components Analysis</td>
</tr>
<tr>
<td>• <strong>Generalized Linear Models</strong></td>
<td></td>
</tr>
<tr>
<td>• Hazard Models</td>
<td></td>
</tr>
<tr>
<td>• Mixed Effect Models</td>
<td></td>
</tr>
<tr>
<td><strong>Non-parametric</strong></td>
<td></td>
</tr>
<tr>
<td>• Neural Networks</td>
<td>• Neural Networks</td>
</tr>
<tr>
<td>• CART (Classification and</td>
<td></td>
</tr>
<tr>
<td>Regression Trees)</td>
<td></td>
</tr>
<tr>
<td>• Random Forests</td>
<td></td>
</tr>
<tr>
<td>• MARS (Multivariate Adaptive Regression Splines)</td>
<td></td>
</tr>
</tbody>
</table>
### GLM as primary method of loss cost analysis

<table>
<thead>
<tr>
<th>Primary method</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalized linear models (GLMs) (e.g., logistic regression)</td>
<td>82%</td>
</tr>
<tr>
<td>One-ways (either loss ratios or adjusted pure premiums)</td>
<td>14%</td>
</tr>
<tr>
<td>Generalized linear mixed models (GLMMs) (fixed and random effects models)</td>
<td>2%</td>
</tr>
<tr>
<td>Machine learning</td>
<td>2%</td>
</tr>
<tr>
<td>Generalized nonlinear models (GNLMs), including GAMs</td>
<td>0%</td>
</tr>
<tr>
<td>Other</td>
<td>0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Additional methods</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Principal component analysis</td>
<td>35%</td>
</tr>
<tr>
<td>Clustering</td>
<td>30%</td>
</tr>
<tr>
<td>Geospatial analysis</td>
<td>30%</td>
</tr>
<tr>
<td>CART/decision trees</td>
<td>21%</td>
</tr>
<tr>
<td>Residual mining</td>
<td>21%</td>
</tr>
<tr>
<td>Machine learning</td>
<td>14%</td>
</tr>
<tr>
<td>Other</td>
<td>5%</td>
</tr>
<tr>
<td>None of these</td>
<td>35%</td>
</tr>
</tbody>
</table>

Source: Towers Watson survey in 2014
## Flexibility of GLMs

Common GLM Applications:

<table>
<thead>
<tr>
<th>Technique</th>
<th>Link Function</th>
<th>Distribution</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classical Regression</td>
<td>Identity: ( g(\mu) = \mu )</td>
<td>Normal</td>
<td>General Scoring Models</td>
</tr>
<tr>
<td>(Ordinary Least Squares)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logistical Regression</td>
<td>Logit: ( g(\mu) = \log[\mu/(1-\mu)] )</td>
<td>Binomial</td>
<td>Binary Target Applications (i.e. Retention)</td>
</tr>
<tr>
<td>Frequency Modeling</td>
<td>Log: ( g(\mu) = \log(\mu) )</td>
<td>Poisson</td>
<td>Count Target Variable Frequency Modeling</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Negative Binomial</td>
<td></td>
</tr>
<tr>
<td>Severity Modeling</td>
<td>Inverse: ( g(\mu) = -1/\mu )</td>
<td>Gamma</td>
<td>Size of claim modeling</td>
</tr>
<tr>
<td>Severity Modeling</td>
<td>Inverse Squared: ( g(\mu) = -1/\mu^2 )</td>
<td>Inverse Gaussian</td>
<td>Size of claim modeling</td>
</tr>
</tbody>
</table>
Other popular techniques and applications

- GLM
- CART (Tree models)
- Neural Networks
- Clustering
- Random Forests
- Time Series/Survival Models

- Product Development
- Marketing
- Underwriting
- Retention
- Claims
- Inforce Management

• Develop product assumptions based on prior products
• Targeted marketing campaigns
• Customer segmentation
• Streamline/reduce UW requirements
• Audit UW process
• Align customer retention with customer value
• Enhance claims forecasting
• Fraud detection
• Refine valuation assumptions
• Understand drivers of policyholder behavior
Predictive modeling workflow

Data
- Import Data
- Explore Data
- Prepare Data

Train the Model
- Known data
- Known responses
- Model

Select Model & Predictors

Use for Prediction
- Model
- Predicted Responses

Measure Accuracy
Best practice #1: Testing and performance monitoring

Modeling Data

Train $X\%$

Test $Y\%$

Validation $Z\%$

Iterative model building

Model Validation

Model Implementation

Ongoing Performance Monitoring
Best practice #2: Refining the model

• Simplify your model when possible
  • Remove predictors, decrease iterations

  ▪ Adding synthetic variables
    – Example: “is the contract owner the insured?”

  ▪ Refining the results with additional models or rules
    – Underwriting rules
    – Principle Components Analysis
    – Combinations of different models
Agenda

- Predictive modeling background
- Case study – Claim settlements forecasting
- Conclusion
Claim Settlements Forecasting

- **Goal:**
  - Improve reserves by producing accurate model to predict insurance claim settlement amounts

- **Approach:**
  - Train a regression model using different techniques
  - Measure accuracy and compare approaches
  - Use model for prediction
Takeaways of case study

- **Interactive environment**
  - Visual tools for exploratory data analysis
  - Easy to evaluate and choose best algorithm
  - Apps available to help you get started
    (e.g., *neural network tool, curve fitting tool*)

- **Multiple algorithms to choose from**
  - Clustering
  - Classification
  - Regression
Agenda

- Predictive modeling background
- Case study – Claim settlements forecasting
- Conclusion
Summary

- Big data and new available technology are creating new opportunities in the insurance industry
- MATLAB can quickly help you get value from your data with its predictive analytics capability
Q&A

Questions!?!