Cutting Edge Tools for Pricing and Underwriting Seminar

Integrating External Data into the Decision Making / Predictive Modeling Process

Casualty Actuarial Society

by Serhat Guven

Fall 2011

TOWERS WATSON

Agenda

- Background

- Data warehousing basics

- External data sources

- Case study: Elasticity Models

2

Background

3

Background

- Goals of predictive modeling

Produce a sensible model that explains recent historical experience and is likely to be predictive of future experience

1. Separate the random components from the systematic components of the estimator

Response Variable = **Systematic Component** + **Unsystematic Component**

Signal: Function of the Rating Factors/Predictors

Noise: Reflects stochastic process

2. Balance predictive power and explanatory effects

Overall mean — **Best Models** — One parameter per observation

Model Complexity (number of parameters)

Underfit: Predictive Poor explanatory power

Overfit: Poor predictive power Explains history

Background

- Goals of predictive modeling

Predict a response variable using a series of explanatory variables

Explanatory variables
Age   Accidents
Limit   Convictions
Region   Credit score

Predictive Model

Parameters Validation Statistics

Response variables
Losses
Claims
Retention

- Larger data storage capabilities and greater access to external data enhances ability to identify the right rate for the risk

---

Background

- Many external factors have been found to be predictive of frequency and/or loss severity. Here are a few for auto…

Safety Features, Annual Mileage, Other Products, Sales Channel, Coverage, Garaged, Model Year, Limits/Deductibles, Territories, # Vehicles, Marital Status, Major Convictions, Driver Training, Credit Score, Occupation, Inexperienced Operator, Detailed Driving Record, Prior Insurance, Comp Claims, Usage, Traffic Density, DWI, Lifestyle Factors, Household, # Renewals, Gender, Minor Convictions, Good Students, Not-at-Fault Accidents, Urban v. Rural, Age, Anti-theft Devices, Make/Model, Method of Payment, Homeownership, # Drivers, Tier

Background

- Modeling is an iterative process
- How does the analyst decide which factors are most valuable?
  - Parameters/standard errors
  - Consistency of patterns over time or random data sets
  - Type III statistical tests (e.g., chi-square tests)
  - Judgment (e.g., do the trends make sense)
- Focus of the section is on gathering data NOT analysis

Review Model

Simplify

Complicate

7

Background

- Greater availability of data requires better understanding of warehousing principles to ensure proper analysis
  - Identifying entity groups
  - Merge strategies

- External data sources for differing entity groups

- Elasticity data set case study
  - Data classes
  - External sources

8

Data Warehousing Basics

9

Data Warehousing Basics

- Basic definitions
  - Primary Key – uniquely identifies the granular construct to be modeled
    — Policy Number
    — Policy Number/Risk
    — Policy Number/Risk/Endorsement
  - Fields – information stored for each granular construct
    — Explanatory factors/ independent variables/predictors
    — Categorizing fields into entity groups

Data Warehousing Basics

- Basic definitions
  - Indexes - important elements in the database
    — Used for key sort and merge elements
    — Primary key is the natural index of data
  - Metrics – measured attributes
    — Dependent variable/Response

- The goal is to identify external data fields that will be merged into the warehouse either on a key or index to better predict the metric

- External data are defined as fields that are not in any internal systems

Data Warehousing Basics

- Proper data management and warehousing store and maintain data in relational tables

- Statistical models inputs require flattened versions of the data
- External data can be defined as additional fields that are merged to specific entity groups on the original data

Data Warehousing Basics

- Standard entity groups that can be used to integrate external data
  - Policy number
  - Named Insured
  - Risk
  - VIN
  - Geographic unit (e.g. census block, zip code, etc)
  - SIC class code
  - Building ID
  - ….

13

Data Warehousing Basics

Two main strategies used to merge external data
- Direct merge on the defined entity group
  - Credit characteristics of named insured
  - Geodemographic elements of the geographic unit
  - Vehicle attributes of the VIN
  - OSHA attributes of an SIC code
  - Competitor's price points

- Indirect merge based on 'proximity' to the entity group
  - Characteristics of the responding fire department
  - Traffic conditions of neighboring zipcodes

14

Data Warehousing Basics

- Issues that will affect the statistical model output
  - Missing data
    — Dummy controls vectors in multivariate models tend to produce biases
  - Changing data
    — External data changes over time (e.g. population growth, credit score definition changes, etc)
    — Need to understand assumptions of data feeding the model and application of the model
    — Concerned about additional prediction error

15

External Data Sources

External Data Sources

Personal

Coverage

Financial

Business

Behavioral

Geographical

External Data Sources

- Personal Lines
- Commercial Lines

Sources

U.S. Census
National Highway Traffic Safety Administration
Experian Mosaic®
Applied Geographic Solutions
Carfax
Highway Loss Data Institute
FireSafe
PerilVision
Personal credit
Commercial credit
Occupational Safety & Health Administration
NCCI/ISO
Safety and Fitness Electronic Records
GPS tracking
U.S. News
Insurequote/Quadrant

Personal Lines Data Sources - Auto

- Personal Lines
  - Auto
  - Homeowners
- Commercial Lines

Sources
U.S. Census
National Highway Traffic Safety Administration
Experian Mosaic®
Applied Geographic Solutions
Carfax
Highway Loss Data Institute
FireSafe
PerilVision
Personal credit
Commercial credit
Occupational Safety & Health Administration
NCCI/ISO
Safety and Fitness Electronic Records
GPS tracking
U.S. News
Insurequote/Quadrant

towerswatson.com

Personal Lines Data Sources - Homeowners

- Personal Lines
  - Auto
  - Homeowners
- Commercial Lines

Sources
U.S. Census
National Highway Traffic Safety Administration
Experian Mosaic®
Applied Geographic Solutions
Carfax
Highway Loss Data Institute
FireSafe
PerilVision
Personal credit
Commercial credit
Occupational Safety & Health Administration
NCCI/ISO
Safety and Fitness Electronic Records
GPS tracking
U.S. News
Insurequote/Quadrant

towerswatson.com

Commercial Lines Data Sources - Workers Comp

- Personal Lines
- Commercial Lines
  - Workers Compensation
  - Commercial Auto
  - Other Commercial

Sources
U.S. Census
National Highway Traffic Safety Administration
Experian Mosaic®
Applied Geographic Solutions
Carfax
Highway Loss Data Institute
FireSafe
PerilVision
Personal credit
Commercial credit
Occupational Safety & Health Administration
NCCI/ISO
Safety and Fitness Electronic Records
GPS tracking
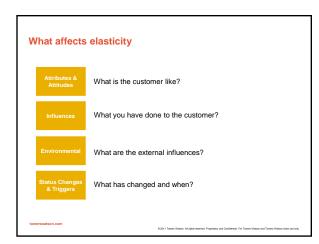U.S. News
Insurequote/Quadrant

towerswatson.com

Commercial Lines Data Sources – Commercial Auto

- Personal Lines
- Commercial Lines
  - Workers Compensation
  - Commercial Auto
  - Other Commercial

Sources
U.S. Census
National Highway Traffic Safety Administration
Experian Mosaic®
Applied Geographic Solutions
Carfax
Highway Loss Data Institute
FireSafe
PerilVision
Personal credit
Commercial credit
Occupational Safety & Health Administration
NCCI/ISO
Safety and Fitness Electronic Records
GPS tracking
U.S. News
Insurequote/Quadrant

Commercial Lines Data Sources – Other Commercial

- Personal Lines
- Commercial Lines
  - Workers Compensation
  - Commercial Auto
  - Other Commercial

Sources
U.S. Census
National Highway Traffic Safety Administration
Experian Mosaic®
Applied Geographic Solutions
Carfax
Highway Loss Data Institute
FireSafe
PerilVision
Personal credit
Commercial credit
Occupational Safety & Health Administration
NCCI/ISO
Safety and Fitness Electronic Records
GPS tracking
U.S. News
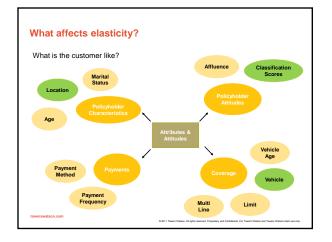Insurequote/Quadrant

Case Study: Elasticity Models

Case Study: Elasticity Models

Data discovery and design is a process
- What affects elasticity
  - General understanding of the process
  - Classifying elements into groups
- Identify external data sources and corresponding timelines
- Construction and manipulation
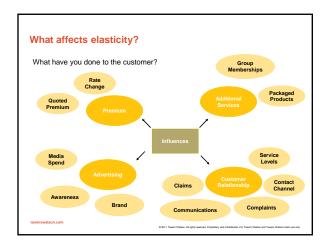- Develop merge procedure

---

What affects elasticity

Attributes & Attitudes	What is the customer like?
Influences	What you have done to the customer?
Environmental	What are the external influences?
Status Changes & Triggers	What has changed and when?

What affects elasticity?

What is the customer like?

What affects elasticity?

What have you done to the customer?

What affects elasticity?

What are the external influences?

What affects elasticity?

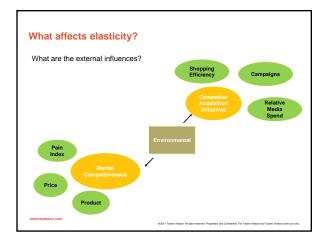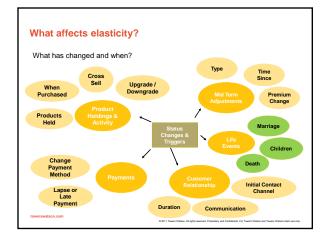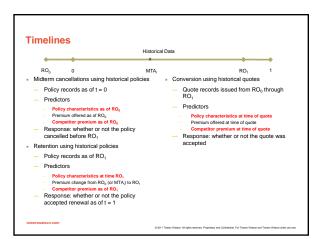What has changed and when?

Identifying External Sources

- Location
  - Census
  - EASI
- Affluence and Classification Scores
  - Third Party Credit Providers (Choicepoint, Experian, L2C, etc)
- Vehicle
  - Polk/Jato/HLDI etc
- Media Spend
- Market Prices
  - Insurequote/Quadrant
  - Indices are constructed via local knowledge

31

Timelines

Historical Data

RO_0 0 MTA_i RO_1 1

- Midterm cancellations using historical policies
  - Policy records as of t = 0
  - Predictors
    - **Policy characteristics as of RO_0**
    - Premium offered as of RO_0
    - **Competitor premium as of RO_0**
  - Response: whether or not the policy cancelled before RO_1
- Retention using historical policies
  - Policy records as of RO_1
  - Predictors
    - **Policy characteristics at time RO_1**
    - Premium change from RO_0 (or MTA_i) to RO_1
    - **Competitor premium as of RO_1**
  - Response: whether or not the policy accepted renewal as of t = 1

- Conversion using historical quotes
  - Quote records issued from RO_0 through RO_1
  - Predictors
    - **Policy characteristics at time of quote**
    - Premium offered at time of quote
    - **Competitor premium at time of quote**
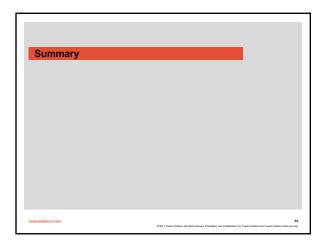 - Response: whether or not the quote was accepted

---

Construction and Manipulation

- Competitor Price Points
  - Policy level analysis
- Market Indices
  - Assessing market conditions
- *Recursive Modeling Builds*

33

Summary

34

Summary

- External data requires understanding of underlying process

- Strong need to identify key drivers of modeling

- Classification creates selection expertise

- Data warehousing provides discipline

35

Contact Details

- Serhat Guven

Towers Watson
Senior Consultant
200 Concord Plaza Suite 420
San Antonio, TX 78216

210.826.2878
serhat.guven@towerswatson.com

36