CENTER for ECONOMIC JUSTICE | FAIR ACCESS FAIR TREATMENT

Use of Big Data in Insurance Products

Presentation to
IRES Career Development Seminar

Birny Birnbaum
Center for Economic Justice

August 8, 2016

The Center for Economic Justice

CEJ is a non-profit consumer advocacy organization dedicated to representing the interests of low-income and minority consumers as a class on economic justice issues. Most of our work is before administrative agencies on insurance, financial services and utility issues.

On the Web: www.cej-online.org

Why CEJ Works on Insurance Issues

Essential Financial Security Tool for Individual and Community Economic Development: CEJ Works to Ensure Access and Fair Prices for These Essential Products and Services, particularly for Low- and Moderate-Income Consumers.

Primary Institution to Promote Loss Prevention and Mitigation: CEJ Works to Ensure Insurance Institutions Maximize Their Role in Efforts to Reduce Loss of Life and Property from Catastrophic Events.

Big Data Defined

Insurers' use of Big Data has transformed the way they do marketing, pricing and claims settlement. Big Data means:

- Massive databases of information about (millions) of individual consumers
- Associated data mining and predictive analytics applied to those data
- Scoring models produced from these analytics.

The scoring models generated by data mining and predictive analytics are algorithms. Algorithms are lines of computer code that rapidly execute decisions based on rules set by programmers or, in the case of machine learning, generated from statistical correlations in massive datasets. With machine learning, the models change automatically.

Personal Consumer Information in Big Data

- Social Media
- Shopping Habits/Purchase History
- Hobbies and Interests
- Demographics/Household Data/Census Data
- Government Records/Property Records
- Web Tracking
- Mainstream Credit Files: Loans, Credit Cards
- Alternative Credit Data: Telecom, Utility, Rent Payment

Examples of Insurer Big Data Algorithms

Pricing:

- Price Optimization/Demand Models
- Customer Value Scores,
- Telematics,
- Credit Scores,
- Criminal History Scores,
- Vehicle Scores,
- FireLine Rating

Claims:

- Fraud Scores,
- Severity Scores

Example: Price Optimization/Consumer Demand Models

Deloitte Presentation at 2014 CAS Ratemaking Seminar

What is the ultimate goal of price optimization? Increase Profit

Insurance pricing can be classified in three levels of sophistication: Basic Rating Plans, Underwriting Models, and Market Demand Models

Market Demand Models: Customer price elasticity to optimize price

A key advantage of including underwriting components is that the insured's price elasticity and demand behavior **is on the final price at the policy level** and not the coverage and sub-component level. Optimizing price on the sub-coverage level creates a gap between the results and the insured's price behavior.

Example: Pricing Models

EagleEye Analytics Real Time Scoring Model

An insurer testimonial: when our underwriter is sitting at his or her desk, and they're looking at a renewal or quoting new business, there are two scores that pop up: a frequency score a kind of underwriting quality score and also a pricing score we use to help price renewal business and new business quoting. The risk scoring goes way beyond just financial data, it uses all of our characteristics in our data whether used for rating or not and then we are also able to bring in third party data, Census Bureau data, to supplement that.

EagleEye Analytics Real Time Scoring Model

Step 3: Build Elasticity Models. The loss ratio model will provide rate indications greater than you will take in one revision. **Elasticity models provide the data you need to forecast how your policies will respond to price change. Talon Elasticity models have multiple segments, not one curve for the entire portfolio.**

Step 4: Optimize. Using the Loss Ration Models, Elasticity Models and incorporating the aging that understands the new business penalty – it's time to make rate level decisions. **Optimization takes these models and applies them specifically to the policies you want. It's a tool to pick the perfect rate level** and forecast future profitability and growth.

Step 5: **Filing Machine Learning models is becoming common**. . . . EagleEye customers have filed and been approved in 45 states across multiple lines of business. . . .**As a bonus – once you switch to machine learning, your competition will not be able to reverse engineer your rating plan.**

Example: Pricing Model

TransUnion Criminal History Score

“TransUnion recently evaluated the predictive power of court record violation data (including criminal and traffic violations)

“While a court record violation is created during the initial citation, the state MVR is updated later and may be delayed depending on a consumer’s response to the citation. For example, if someone pleads guilty and pays a ticket immediately, the state MVR will be updated in approximately two months. If the ticket is disputed in court, in contrast, the state MVR may not be updated for 6–19 months or longer.

“Also, as court records are created when the initial citation is issued, they provide insight into violations beyond those that ultimately end up on the MVR—such as violation dismissals, violation downgrades, and pre-adjudicated or open tickets.”

Example: Claim Fraud Scores, Claim Severity Scores

LexisNexis Claim Tools

“LexisNexis (LN) seeks to provide a Single Point of Entry for delivering all of information directly back into a carrier’s system whether from a marketing standpoint, underwriting process or especially the claims part.

“LN has over 10,000 data sources that feed into its infrastructure each month and has contributed information from the industry.

“Claims Data Fill” – deliver data and analytics directly into claims system in the claims process regarding parties, vehicles and carrier information. Used to verify information provided to insurers and provide indicators beyond the data to identify whether a social security number is an indicator of fraud or whether an address provided is a good address. Has an analytic component at first notice of loss and throughout the claim, constantly monitoring the claim looking for fraudulent activities. Real time data verification and enhancement with fraud scoring and attributes

LexisNexis Claim Tools (con't)

"Example, insured calls in, rear-ended, all I got was license plate:

"Claims Data Fill takes that license plate, reach out to DMV to get vehicle registration to get VIN number, we have policy database and get the carrier and policy information, take the registered owner, go out to public records, pull back their address, date of birth, telephone number, social security, wrap that into a package and put it back into our system, 88% of the time done in less than 5 seconds.

"Take minimum information provided at first notice of loss, provide a fraud score at the initial notice of loss. Daily monitoring of claim every time new information comes in, able to run various scores: fraud scores, severity score."

Example: Fraud Scores

LexisNexis: “Severity Focus”

“Identify claims with the potential to become severe:
SeverityFocus utilizes advanced predictive modeling to identify claims with the potential to become severe as they develop claims that otherwise would go undetected until much later.

“SeverityFocus does not constitute a "consumer report" as that term is defined in the federal Fair Credit Reporting Act, 15 USC 1681 et seq. (FCRA).”

Example: Claims Scoring

StatSoft's Predictive Claims Flow™

"A predictive analytics and reporting solution for property and casualty insurance companies, can help you reduce loss ratios and improve bottom-line profitability, often within a few months of implementation. StatSoft's Predictive Claims Flow™ solution incorporates predictive modeling at every stage of an insurance claim. This closed loop system has a unique scoring system that rates each claim at its inception on its propensity for fraud and then continually rescores the claim as it goes through each step of a claim's lifecycle."

Example: Fraud Scores

Infosys Social Network Analysis

"The SNA tool combines a hybrid approach of analytical methods. The hybrid approach includes organizational business rules, statistical methods, pattern analysis, and network linkage analysis to really uncover large of amounts of data to show relationships via links. When one looks for fraud in a link analysis, one looks for clusters and how these clusters link to other clusters. Public records such as judgments, foreclosures, criminal records, address change frequency and bankruptcies are all data sources that can be integrated into a model. Using the hybrid approach, the insurer can rate these claims. If the rating is high, it indicates the claim is fraudulent."

Example: Fraud Scores

Infosys: Social Customer Relationship Management

"Social CRM is neither a platform nor a technology, but rather, a process. It is important that insurance companies link social media to their CRM. Social CRM . . . gathers data from various social media platforms. It uses a "listening" tool to extract data from social chatter,. . . .The reference data along with information stored in the CRM is fed into a case management system. The case management system then analyzes the information based on the organization's business rules and sends a response. The response, from the claim management system as to whether the claim is fraudulent or not, is then confirmed by investigations independently, since the output of the social analytics is just an indicator and should not be taken as the final reason to reject a claim."

Big Data and Modeling of Rates/Prices/Claims/Customer Value

<u>Old Old School Big Data</u>: Advisory Organization Loss Costs. Oversight of Data, Advisory Organization, Analytic Techniques, Filings, Complete Transparency

<u>Old School Big Data</u>: Credit-Based Insurance Scores. Limited Consumer Protections for Completeness and Accuracy of Data via the FCRA, Limited Oversight of Modelers and Models, Limited Transparency

<u>New School Big Data</u>: Predictive Modeling of Any Database of Personal Consumer Information. No Consumer Protections for Completeness and Accuracy of Data, No Oversight of Modelers and Models, No Transparency to Consumers

Insurer Use of PO and Big Data Scoring Models Lack Fundamental Consumer Protections

- Accuracy and Completeness of Data
- Oversight of Data Bases
- Disclosures to Consumer About Data Used, How Used and Privacy Protections
- Consumer Ability to Challenge False Information
- Regulators' Knowledge Of and Capability to Provide meaningful Oversight
- Prevent discrimination Against Low-Income and Minority Consumers and other protected classes
- Asymmetric Use of Data
- Greater Cybersecurity Danger for Consumers and Insurers

The Way Forward: 4 Steps

1. Identify What Insurers are Doing
2. Monitor Market Outcomes
3. Create an NAIC Resource for the States for Big Data Analytics
4. Develop a 21^{st} Century Approach to Oversight of Risk Classifications

Identify What Insurers Are Doing

To a great extent, regulators – and, of course, consumers and policy makers – do not know what types of information insurers are using and what they are using the information for and how they are using it.

A logical first step is to develop a template for states to use, with assistance from the NAIC for collection of requested information, to request from insurers the sources and uses of data for various insurance functions. For each source of data, the insurer would provide a name/description of the data, the source of the data and the use or uses of the data -- pricing (including underwriting), marketing, claims settlement, antifraud and other.

This periodic survey will provide regulators with the basic overview of what types of information are being used by insurers and what the information is being used for. This information is essential for regulators to respond to policy makers and to foster public discussion over potentially controversial types of data.

Monitor Market Outcomes

The old regulatory model of monitoring everything that goes into insurer marketing, pricing and claims settlement practices and models is not feasible in the era of Big Data. Regulatory Big Data is needed. The data needed for a robust market analysis – one that includes the ability to monitor the affordability and accessibility of insurance in underserved areas as well the ability to perform enhanced market analysis to focus regulatory resources on problem companies and problem markets – is transaction data on premium quotes, policies issued and claims. Insurance regulators lag other financial regulators in the ability to monitor market outcomes and the collection of granular data for market monitoring.

Create an NAIC Resource for the States for Big Data Analytics

Just as the NAIC provides resources to assist the states in other areas – technical/actuarial capability at the NAIC to assist the states with PBR; collection and compilation of massive amounts of financial statement data to assist with financial analysis, collection and compilation of Market Conduct Annual Statement data to assist states with market analysis, to name just a few examples – the NAIC should develop resources to assist states in analyzing new big data models used by insurers.

The NAIC resource would not be a regulator and would not provide regulatory opinions. Rather, the NAIC resource would provide states with technical – actuarial and statistical – expertise to answer states' questions about a big data pricing or claims model. The NAIC resource would also assist states in accepting and processing large data sets as part of analysis of a pricing or claims model.

Modernize Regulatory Oversight of Risk Classifications

Risk classification represents insurers' – and society's – decisions about how to group consumers for the purpose of assigning premium. Risk classification determines the affordability and availability of essential financial security tools – insurance – for consumers. In the vast majority of states, the only justification needed for a risk classification is a correlation. But, in an era of Big Data and the modeling of rates and risk classifications, such a test is arbitrary and opaque. We see this in the huge differences in rate impact of the same risk classification across insurers and even for an insurer across states. There is a need for a 21st century approach to oversight of risk classifications – more transparency and more accountability