The New Paradigm of Property & Casualty Insurance Pricing:
Multivariate analysis and Predictive Modeling

The ability to effectively price personal lines insurance policies – to accurately match rate with risk - is arguably the most important competency required of Canadian Property and Casualty insurers. Pricing each risk with surgical precision can provide substantial competitive advantage and long term profitability. Over the past decade, a number of insurers have emerged as leaders in pricing sophistication – Progressive and GEICO in the U.S., Intact in Canada, and many EU insurers. Due in part to advances in computing power, but more so due to highly competitive market forces that drive innovation, these companies have complemented conventional actuarial methods with data mining and analytical techniques to produce more stable and accurate rating structures.

These techniques, commonly referred-to as "multivariate analysis" (MVA) - have long been an essential part of the direct marketer's tool kit, are now being adopted by Canadian insurers and are on the verge becoming part of everyday business practices. As Canadian insurers employing these techniques become more sophisticated in their use and application, the performance divide between these innovators and their competitors will grow.

Current P&C Industry Pricing

Company Actuaries create rating structures that generate a premium for each risk in a given portfolio. For personal auto insurance this process usually includes:

- Analyses of various policyholder type characteristics and their impact on claim risk (i.e. claim frequency and severity),
- Industry reports and tables that reflect the most recent industry trends regarding claim risk; and
- Corporate financial objectives and competitive market factors,

Industry Actuaries are challenged with bridging the divide created by analytical/statistical indications and provincial regulatory requirements – two conditions that are often at odds with each other in a very competitive environment.

The conventional approach to rate setting employed by most insurers today dates back to the 1960's when data processing capabilities were limited. Historically, rating analyses are conducted on a *univariate* basis; in other words, they look at how changes in one characteristic result in differences in claim frequency or severity. Loss frequency and severity measures are determined for commonly-used rating characteristics like "Claim history" (it's well established that the longer it has been since a driver has incurred a loss, the less likely they are to incur one in the future) and "Vehicle use" (a vehicle driven for personal use is less likely to incur a loss than a vehicle used for business; presumably because the vehicle is simply on the road less).

Most industry practitioners use the Driving Record (the period of time a driver is "claim free") variable to establish "base" rates for each vehicle. Generally, Driving Record categories range from 01 to 06 (usually up to "06" or six years claims free; but with better and more reliable data capture some companies have introduced 10, 15 and 20 year driving record groups), and every risk starts off with the same "base rate" set for the driving record group within which they fall. Then another rating variable, let's say Driving Class (type of use for the vehicle – pleasure only, drive to work, business use, etc), is applied. To determine how the base rate for each Driving Record group will change with the application of the next rating factor, a *differential* is applied. The differential value is a measure of the difference in observed (actual) claim losses for each

different Driving Class compared to the overall average. I've tried to simplify the process In the following example for Collision coverage. You can see how the premium charged for the vehicle changes as the different rating characteristics are combined:

See Table 1 below for example:

Table 1

Driving Record	01		02		03		04		05		06	
% of Drivers	3%		4%		8%		10%		15%		60%	
Loss Rate	5.5%		4.8%		3.1%		1.5%		1.1%		0.6%	
Differential	429.7%		375.0%		242.2%		117.2%		85.9%		46.9%	
Total Collision Premium	$ 10,313		$ 12,000		$ 15,500		$ 9,375		$ 10,313		$ 22,500	
Premium/Driver	$ 344		$ 300		$ 194		$ 94		$ 69		$ 38	
Driving Class	01	02	01	02	01	02	01	02	01	02	01	02
% of Drivers	80%	20%	80%	20%	80%	20%	80%	20%	80%	20%	80%	20%
Loss Rate	1.10%	2.00%	1.10%	2.00%	1.10%	2.00%	1.10%	2.00%	1.10%	2.00%	1.10%	2.00%
Differential	0.86	1.56	0.86	1.56	0.86	1.56	0.86	1.56	0.86	1.56	0.86	1.56
Total Collision Premium	$ 7,090	$3,223	$ 8,250	$ 3,750	$ 10,656	$ 4,844	$ 6,445	$ 2,930	$ 7,090	$ 3,223	$ 15,469	$ 7,031
Premium/Driver	$ 295	$ 537	$ 258	$ 469	$ 167	$ 303	$ 81	$ 146	$ 59	$ 107	$ 32	$ 59
-												
-												
-												
etc.												
Key Assumptions												
Avg Collision Loss Rate	1.28%											
# of Drivers	1,000											
# of Claims	13											
Total Collision Losses	$50,000											
Loss per Claim	$ 3,846											
Target Loss Ratio	62.5%											
Required Premium	$ 80,000											

In this simplified example, there are only two Driving Class groups – 01 (Personal use only) and 02 (Business use only). As each rating variable is added (e.g. Years Licensed, FSA location, Vehicle Rate Group, etc.) differentials for the variable are used to repeatedly adjust the vehicle premium to reflect the risk represented by the values of each rating characteristic or factor used by the insurer to determine the premium for the risk. Each risk in a portfolio falls into a small group/profile that best reflects the losses anticipated by that profile.

It is the use of these cross tab type reports that determine the variation of premium between groups of policy holders with different characteristics. As more factors or variables are introduced into the cross tab reports, we observe more groups and a lesser number of policyholders within each group. In effect, the analysis becomes more granular, resulting in many different possible premium values across the policyholder base. This increase in the number of risk groups or categories results in increased granularity, greater discrimination between risks and more accurate pricing.

Ultimately however, using this methodology the number of different possible premium levels will be restricted by the number of groups that are included in the analysis.

Applying MVA and Predictive Modeling Techniques to Home and Auto Insurance
Multivariate analytical techniques, made possible by advances in computing, take a different approach to predicting risk by:

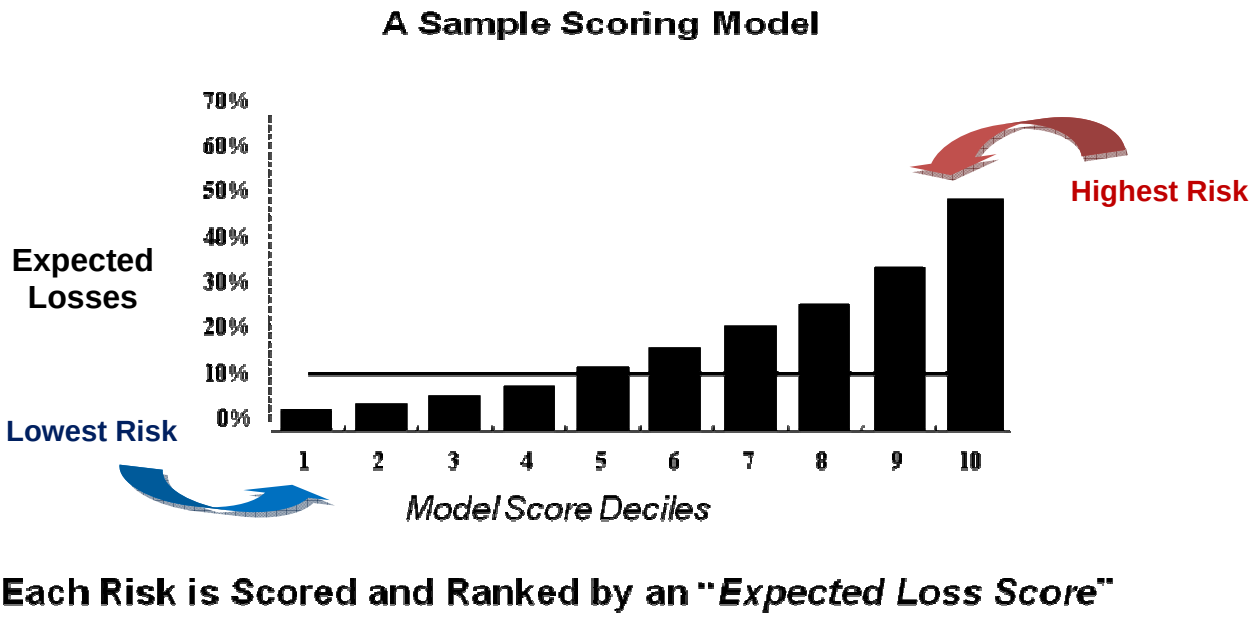
- focusing on individual level data, so the estimate of risk is more *granular* and,
- taking into account the effects (*interactions)* that many different characteristics (variables) of a risk have on one another (thus the term "multivariate" approach vs the "univariate" approach typically employed by most insurers today)

These techniques have been mastered by data miners across academic (e.g. scientific research) and business (e.g. marketing and risk management) practices as way to predict behaviour. Publishers use Predictive Modeling (a form of multivariate analysis) to create measures of the likelihood that a customer will purchase a particular product (e.g. magazine subscriptions); Banks use these tools to create measures (e.g. credit scores) of whether a client will be able to meet lending obligations for a loan or mortgage. Google uses these techniques to predict internet search behaviour.

Similarly, P&C insurers can use predictive models to predict claim behavior. Essentially, predictive models

- identify the characteristics that best predict risk;
- produce a scoring equation that can be maintained and updated;
- calculate a score that represents the expected losses for each risk in the portfolio;

Chart 1

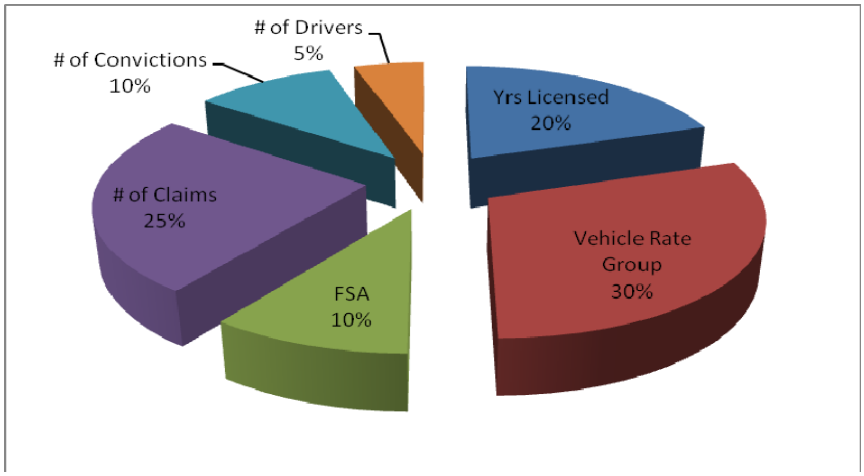


The Scoring model or mathematical equation is comprised of the fewest number of characteristics (variables) possible (usually 10 to 15), each of which are responsible for a discrete "amount" of the expected loss behavior that when added together account for 100% of the expected losses for the risk.

A simple way of looking at this approach is to consider each characteristic in a model equation like a different piece of a pie where the pie represents the total "expected loss" amount. In Chart 2 below, the pieces (characteristics of a auto insurance risk identified by the model) fit

together to form a whole, perfectly shaped pie; in other words, the set of variables in the equation are the optimal combination that provides the most accurate measure of the expected losses for the risk. Whereas it may appear that by adding more information (ie. another piece to the pie) one could generate an even more accurate assessment of the risk, the statistical processes account for the relationships between different characteristics such that the addition of another characteristic to the pie would only distort its shape, and not add any additional value in accurately measuring risk.

Chart 2



It is our experience that multivariate or predictive modeling techniques provide even more granularity and better understanding of the differences between individual risks than conventional approaches used by most insurers today. This is achieved as risk measures produced by predictive modeling outcomes are produced for each individual risk rather than by group differentials. In essence, the techniques enable more accurate matching of rate with risk. Those relatively few insurers in Canada employing these methods are able to acquire business by offering lower rates for risks that the general market is overpricing, and have higher prices to avoid taking on risks that the general market is underpricing. The overall benefits of this result include:

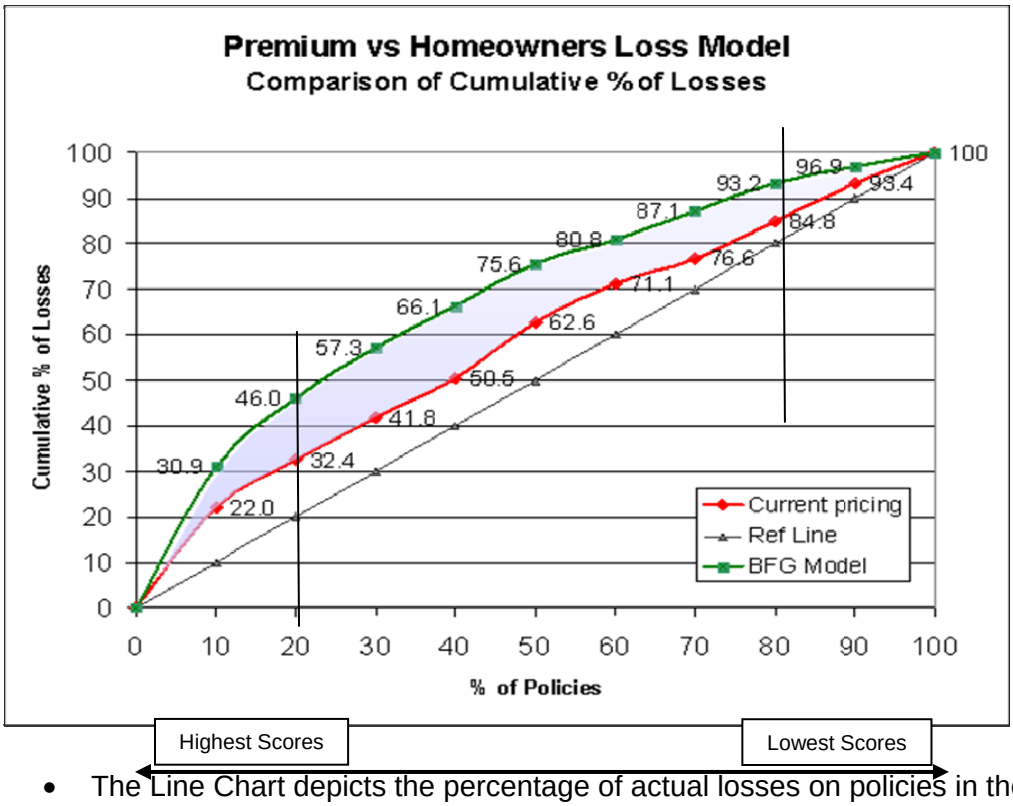
- Better risk selection and pricing
- Reduced underwriting expenses
- Improved underwriting results

Results from Case Studies

Whether the application of multivariate analytical techniques can improve upon the rating accuracy is ultimately determined by how much better these measures predict losses compared to the premium assigned to a risk (vehicle or property) using conventional rate-setting paradigm. Presented below is an example of actual observed results for a portfolio of homeowner's policies:

- A "Homeowners" Claims Risk Scoring model was produced for a Canadian insurer
- Each policy in the portfolio was scored at its effective date. The score represented the expected losses on the policy in the policy year.
- Each policy was ranked from highest risk (score) to lowest risk (score).

Chart 2 – Homeowners Property Insurance Model



Highest Scores ⟷ Lowest Scores

- The Line Chart depicts the percentage of actual losses on policies in the portfolio that occurred after the policy effective date
  - Based on the <u>model's prediction of losses</u>, policies with the highest 20% of scores generated <u>46%</u> of total losses while those policies in the lowest 20% of scores accounted for only <u>6.8%</u> of losses
  - Based on the <u>premium</u> charged for each policy, the policies with the highest 20% of premium produced 32% of all losses and those policies with the lowest 20% of premium produced 15.2% of losses

- The shaded area represents the "lift" or increased accuracy in loss prediction produced by the model over the insurer's current rating structure, essentially representing losses *not-accounted-for by existing rates.*

Conclusion

The application of data mining tools and multivariate modeling techniques can substantially improve current rating structures for Property and Casualty insurers. Because of this capability multivariate modeling techniques are gaining traction and becoming an industry standard. The primary challenge is to effectively action these tools; Insurers must become more familiar with these techniques and adopt them as a daily part of doing business. “Data Miner’s” familiar with these tools can provide critical assistance to company Actuaries in this regard.

This incremental level of accuracy in predicting losses enables insurers to price policies more accurately than competitors, improve portfolio profitability and provide substantial long term competitive advantage.