The Future of Predictive Modeling

Man Versus Machine?

By Peter Lee and Serhat Guven

Predictive models are now used by insurers around the world to run their businesses. While some believe they will become automated commodities, others say they make a true difference only when paired with human judgment and expertise.

Who remembers the concept of the paperless office? The development and widespread deployment of powerful desktop computers toward the end of the 20th century was widely heralded as the end of the need for paper in many workplaces. The term conjured up a vision of the efficiency and dominance of the machine.

Fast-forward a couple of decades, and who could legitimately argue that this has taken place?

There is no disputing that technology has changed the way we work. Certainly, people have grown to rely on their laptops, PDAs and smart phones, but the human instinct to make things simple still holds true. The result is more a case of man working with machine.

The parallels are worth remembering when looking at what the future holds for insurance applications of predictive modeling.

The Story So Far

Predictive models use historic data to identify and quantify patterns and trends that can be used to predict future behavior. So, taking a simple example from household insurance, an area prone to flooding in the past can reasonably be expected to carry a higher risk of future flood damage, which should be reflected in the price charged.

In many developed markets, predictive models are now the de facto standard for setting prices, particularly in personal lines such as auto and homeowners insurance, where the volumes of data are substantial. With advances in the availability of internal and external data, they have also become common in commercial insurance.

An important step in this process was the practical implementation of generalized linear models (GLMs). As these were multivariate models, insurers could look at the impact of a number of factors simultaneously and isolate the "pure" effect of adding or taking away individual factors in building a technical price.

As techniques developed further, analysis was not limited to just predicting the effect of factors such as age and geography on claim costs. Data combined with the increasingly available range of customer and lifestyle information made it possible to analyze a wider range of customer behaviors to help determine the final retail price and also to enhance customer segmentation. Taking this further, and used in tandem with specific optimization algorithms, some insurers are now quoting prices that are tailored to the individual — in some cases, in real time. What's more, companies are taking pricing techniques and applying them to underwriting, fraud mitigation, claim handling and retention management.

"Companies are taking pricing techniques and applying them to underwriting, fraud mitigation, claim handling and retention management."

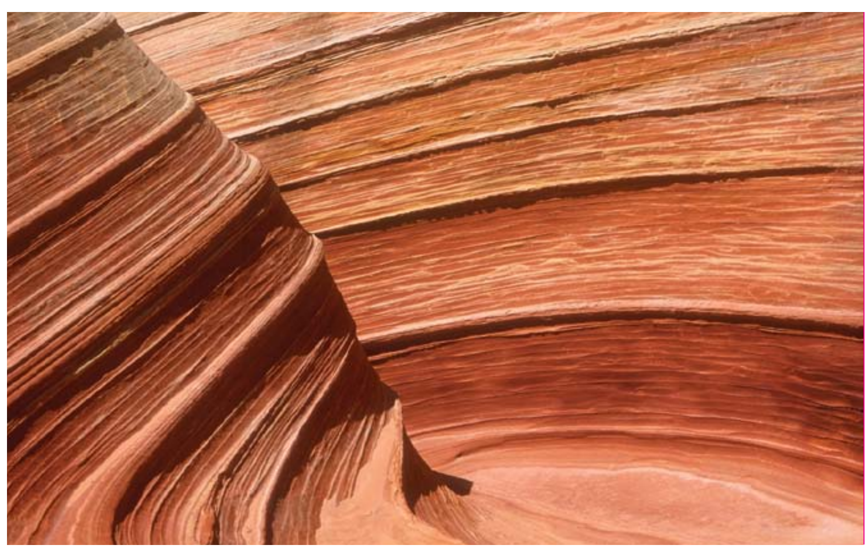
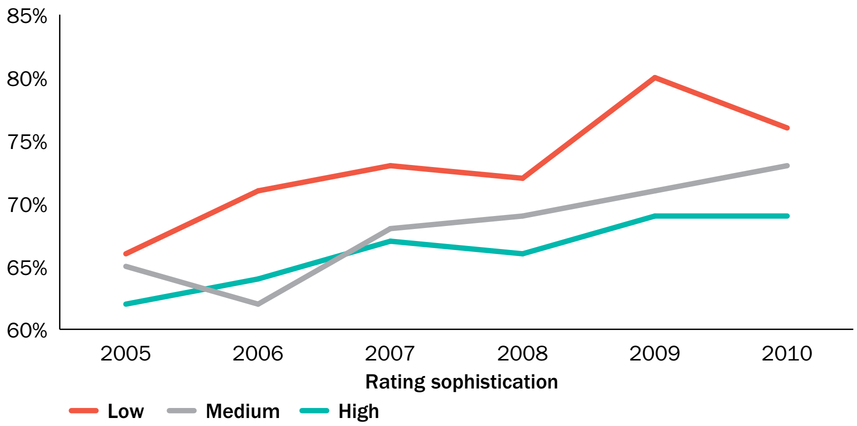


Figure 1. U.S. personal auto pricing sophistication versus performance



*Loss ratios include ALAE.
Source: Towers Watson 2011 Predictive Modeling Survey

Predictive modeling has improved business performance worldwide. *Figure 1,* for example, shows the relative performance of U.S. auto insurers by rating sophistication.

"Technology makes many things possible, but execution and implementation remain important."

Competition Rules

In spite of these impressive results, the reality is that prices based on models are just estimates of the amount needed to cover costs and build in some profit. As Niels Bohr, a Nobel Laureate in physics once said: "Prediction is very difficult, especially if it's about the future."

Insurers have little to gain from attempting to build a perfect model under laboratory-type conditions. The rules of costs versus benefits apply; the focus has been and should remain on meeting business needs and, ideally, improving a company's relative position in the market.

In this respect, the most significant "push" factor by far for predictive models is increased competition. As competition increases, the penalties for pricing errors increase, either from failing to win or retain business because prices are too high or from underpricing and, as a result, attracting a block of unprofitable customers.

Nowhere has this been seen more vividly in action than in markets where price transparency has increased as a result of a lot of business moving online. In the U.K., for example, price comparison websites, which enable consumers to input their personal information and compare quotes from a large number of providers simultaneously, now account directly or indirectly for more than 50% of auto policies sold.

Research shows that the majority of policies bought through this channel appear in the top three to five cheapest quotes received. With the environment suddenly more competitive, insurers had to address previously undetected issues in their predictive models of risk premium underpinning their prices. Unfortunately, these weaknesses were not detected and addressed before many players incurred significant underwriting losses.

The Pull of Technology

Technology helps insurers creating models respond to pressures, including competition. Increased computing power and software improvements have made it possible for innovative insurers to respond with more sophisticated pricing strategies and to build more predictive models. Recent advances in quotation systems have also made it easier to offer individual customers the right products as well as optimized prices.

Technology makes many things possible, but execution and implementation remain important. Depending on the speed and dynamism of a market (U.K. auto would be an example of a very dynamic market), more sophisticated pricing capability needs to be linked to enhanced management information. Model predictions need to be compared with reality and actual performance monitored. One does not black out a pilot's windows just because airplanes can be flown by computers.

The Internet, and wider wireless connectivity, in general, have created a potential data boom that could improve models and insights, and increase the ability to predict or even introduce completely new products. One new way of evaluating risk is usage-based auto insurance (UBI), which collects and processes entirely different risk data such as driving behavior. Operating losses and very tight margins have driven these efforts. Even so, there is still the question of whether the automation of predictive models is the natural evolution of this technological revolution.

Method Versus Techniques

We need to distinguish between method and techniques — and turn to baseball — to explain why the future is more nuanced than an automated predictive modeling process.

The 2011 movie *Moneyball,* starring Brad Pitt, is based on Michael Lewis's book, which tells the story of the 2002 Oakland Athletics baseball team. The club had just lost a number of star players to wealthier teams and did not have the money to bring in comparable replacements. In an unconventional effort to help the A's compete, the general manager of the team, Billy Beane, hired a junior statistician to analyze statistics of players and free agents from across the major and minor leagues. The statistician used predictive analytics to identify undervalued players to bring onto the team.

The results were immediate, with the A's winning the American League West and setting a league record of 20 consecutive wins.

The models were not perfect — they didn't win the World Series, after all — but they did propel the team ahead of many of its opponents. What was more important was that the models were not built in isolation by the statistician. Rather, it was a perfect illustration of man working with machine. The baseball insider used his knowledge to define the variables (metrics) of the model, but statistics, not subjective judgment, was used to rank prospects.

Insurers can learn a valuable lesson — techniques alone do not represent the method. Without applying the right techniques to the problem and judgment from an understanding of the underlying business processes, a lot of the data and results that are produced from technological advances are in danger of being worthless, or worse still, misleading.

The Data Explosion and Data Mining

In their raw form, the majority of data are "dirty" and unstructured. For instance, data culled from UBI programs are often subject to signal fallouts or false signals. The data have to be cleaned and appropriately structured before they can begin to help assess risk.

What's more, any data errors that find their way into a model through an automated process are likely to be replicated in the holdout sample that machine-learning techniques use to cross-validate results, leading to incorrectly concluding that a variable is predictive of the future. Further advances in computing power may improve interpretations but are unlikely to eradicate misinterpretations. Pure machine-learning and artificial-intelligence-based approaches are implemented only in very controlled environments where the penalty from a prediction error is small.

Data mining is a catchall term for organizing and extracting patterns from data. The power of appropriate mining techniques for data discovery and identifying relationships among data points is undeniable, but the question of how to quantify relationships in a practical way within the mining exercise remains.

So we believe that exploratory data mining is most useful for predictive models as a way to initially clean and filter data that are relatively new to companies or need to be broken down into more manageable and interpretable parcels. This might apply to sources such as web-click stream data (where explanatory factors are not obvious) or to a post/ZIP-code analysis (which could generate as many as 300 variables for further analysis). This phase of the modeling process is often referred to as exploratory analysis, and for new data sources, this is where raw data fields become potential explanatory variables. Even with new data sources, the human expert will still be able to use domain knowledge to improve on the initial variables uncovered by the exploratory mining techniques.

What's more, good analysis distinguishes between what statisticians refer to as signal and noise. Understanding the process behind the modeling objective means that models can be framed to maximize the predictive information (signal).

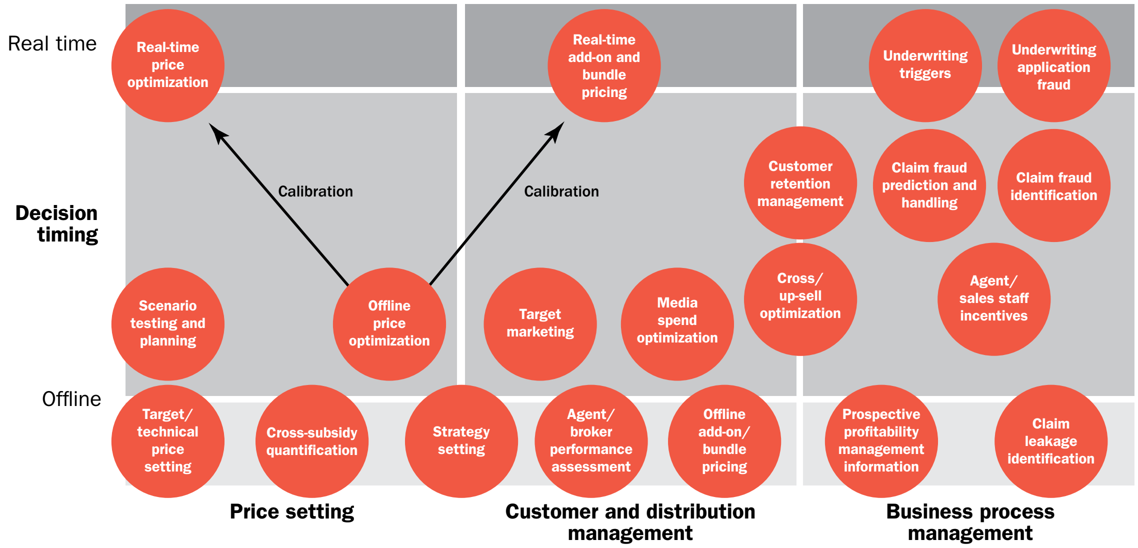
Peter Lee
Specializes in P&C pricing and product management.
Towers Watson,
Epsom, U.K.

Serhat Guven
Specializes in P&C pricing and product management.
Towers Watson,
San Antonio

Recommended Predictive Modeling Core Requirements and Features

- Clear and transparent framework:
  - Underwriters and analysts need to be able to understand the models
- Opportunity for using business knowledge in the process
- *Not* a best statistical fit to historic data:
  - Test predictiveness and consistency of patterns over time and across validation samples
  - Do solutions pass the "does it make business sense" test?
- Data are appropriate for modeling:
  - Quality, scope and quantity suitable to meet objectives
  - Structured in line with underlying processes being modeled to maximize the signal

Figure 2. Applications of predictive modeling



"Good analysis distinguishes between what statisticians refer to as signal and noise."

Simply replicating a signal in past data dooms the user to repeat the past — a good model is a balance between past experience and the company's expectation of future trends. A modeling framework that allows human experts to enrich the process with their knowledge and experience is required. GLMs are time-tested and still provide a robust framework that can also easily incorporate new insights from a range of alternative methods. For example, Towers Watson has developed innovations that complement and enhance GLMs, such as automated identification of potential multifactor effects and cross-validation-based approaches that reduce the replication of random noise.

Domain knowledge has also helped insurance actuaries implement solutions such as spatial smoothing in geographic and vehicle classification analyses when data are sparse.

Realizing the Benefits

Predictive models alone do not create business value, but rather need to be effectively deployed either into a decision-making or business process. *Figure 2* illustrates the diversity of potential predictive modeling applications — the uses become more sophisticated, spreading from price setting and improving marketing effectiveness to creating system rules with parameters that improve wider business processes. Depending on the sophistication and level of competition in the market in which an insurer is operating, some will be more appropriate than others. But whatever a company's objectives, benefits will only be realized when appropriate techniques and effective implementation are deployed.

Data, technology and predictive modeling have their own roles but are really the beginning and not the end solution. They must also be part of a larger overall strategy, advancing product design, branding and distribution channels.

Man With Machine

The power of technology is not simply alchemy that transforms the way we do business, as the theory of a paperless office proves. The coexistence of man working with machine typically produces the best results. Insurers that want predictive models to improve their competitiveness would do well to keep this in mind.

Undoubtedly, as the range and volume of data collected increase, we will need more analytical techniques, particularly in the exploratory analysis phase. Unfortunately though, there are no shortcuts when it comes to having a clear strategy for how to use models or for how to implement them. Expert interpretation of the patterns quantified and knowledge of the issues at hand are essential for success.

For comments or questions, call or e-mail Peter Lee at +44 1372 751079, peter.lee@towerswatson.com; or Serhat Guven at +1 210 767 3811, serhat.guven@towerswatson.com.