THE LAST THING A FISH NOTICES IS THE WATER IN WHICH IT SWIMS

COMPETITIVE MARKET ANALYSIS: AN EXAMPLE FOR MOTOR INSURANCE

Topic:

PRICING RISK

Authors:

Santoni, Alessandro
Towers Perrin — Via Boezio, 6 00193 Rome, Italy
Tel. +39 335 7497624 – Fax +39 06 32803501
alessandro.santoni@towersperrin.com

Folch, Eduard
Towers Perrin — Suero de Quinones, 40-42, 28002 Madrid, Spain
Tel. +34 91 5903009 – Fax +34 91 5903081
eduard.folch@towersperrin.com

Sanche, Bob
Towers Perrin — 175 Powder Forest Drive, Weatogue, CT, U.S.A.
Tel +1 860 843 7123 – Fax +1 860 843 7001
bob.sanche@towersperrin.com

Abstract:

This paper describes a statistical model used to assess an insurance company's competitive position and how to use the result of the model to improve an insurer's pricing strategy. We call these techniques/methodologies: Competitive Market Analysis (CMA), which we define as an analytic process through which premiums charged in the market for a very large number of risks are used as an input to a model that can answer the following questions:

- How can the market's competitive intensity be summarized in one straightforward measure?
- Which are the market segments with the highest/lowest competitive intensity and what is my positioning for customers falling into these segments?
- What is my average overall premium compared to my competitors?
- In which segments am I particularly expensive or inexpensive?
- Who is competing against me in these expensive, or inexpensive, segments?

Finally, these results are combined with other analytics to optimize an insurer's pricing strategy. We also provide a brief overview of two other critical components of the rating processes: costing and analysis of elasticity; however, our main focus is on CMA.

Keywords: Decision tree, Clustering algorithms, Competitive Market Analysis, Motor rating.

INTRODUCTION

Insurers have traditionally used a technical pricing approach (costing) in personal lines. Costing can be used strategically to identify profitable market niches not yet identified by competitors, and it's particularly important in countries where customers have a high propensity to "shop" because they will gravitate towards insurers who underprice the risk. But costing alone is no longer enough. Sophisticated insurance companies are currently developing a more comprehensive analytic approach to their pricing strategy. To get the larger picture, insurers are also assessing:

- competitive position
- market dynamics
- regulatory and distribution channel constraints
- customer behavior and lifetime events.

Insurance companies should want to optimize pricing to meet long-term profitability and growth objectives. While several insurance companies in various countries have now successfully used these different pricing inputs to enhance profitability and grow market share, many insurers are still assessing their competitive positioning by simply comparing a few insurance quotes through the Web or rating manuals filed with regulators. Given the growth in rating plans' complexity and the tendency to focus on particular segments and market niches, we have found this approach particularly misleading.

This document describes a comprehensive and structured process for assessing an insurer's competitive position and shows how these analyses can be implemented in pricing strategy. While we focus primarily on CMA analyses, we also provide a quick overview of costing and assessing customer behavior (elasticity of demand). There are other critical parts to proper price optimization, including lifetime customer value concepts that incorporate potential profitability from the sale of other insurance products, but because our focus is on CMA, they are not discussed in this paper.

The paper has three parts:

- Description of the rating process elements: cost pricing, competitive pricing and elasticity of demand
- The decision tree techniques and, in particular, Classification Tree Analysis
- CMA and the use of classification trees for understanding competitive intensity, using a motor insurance example of an application of classification trees.

RATING PROCESS

The price that an insurer should charge for a given risk is a point in the matrix defined by the following axes:

- **Insurance cost:** This includes claims, expenses, investment income and the cost of capital. This axis measures profitability.
- **Competitiveness:** This axis measures market positioning by customer and the intensity of competition (how much insurance companies fight to gain market share in particular niches.)
- **Customer behavior:** This axis measures the elasticity of demand for a company and its existing/new customers (elasticity of demand in and out) by customer segments.

Most insurers implement their rating strategy using only the insurance cost axis. However, there is a growing trend for personal lines insurers to either consider all of these aspects separately or to combine them in an integrated approach through an algorithm that optimizes the rating plan taking into account cost, elasticity and competitiveness. This process, called "Price Optimization Management," consists of an algorithm that, for a given objective function (for example, maximization of profits or return on capital) and given a number of constraints (e.g., cost, elasticity, intensity of competition), provides a number of rating structure alternatives (often represented on an efficient frontier).

The three components of the pricing process will have different priorities for companies with different objectives, or that are in different markets or different stages of the insurance cycle. In unprofitable markets, an insurer will use cost pricing to minimize cross-subsidies in its rating structure, and will worry less about shifts in portfolio mix that will generate losses regardless of its relative price. If an insurer wishes to grow, it will be mainly interested in its competitive position and in understanding elasticity of demand inwards (conversion rates). If an insurer is defending its book of business in a profitable and particularly aggressive market, its priority will be on retention rates.

Below we briefly analyze the three components of the rating process as described above.

Cost pricing

Cost pricing is the oldest and most common part of the rating process. Most companies and their actuaries have devoted significant time and technical expertise to analyzing the pure risk premium. In particular, personal lines insurers have put a lot of effort into setting prices by segment: age, driver history, car type, geographic location, and so on. They are also investigating more innovative factors to help reflect the interactions between variables so they can avoid cross-subsidies and target specific market segments. This approach takes into consideration only the risk premium component of the economic cost of underwriting that includes expenses and cost of capital.

The risk premium is, in most cases, only 60% to 70% of the equilibrium premium. (To arrive at the tariff premium, we need a loading for profit.) The remainder includes expenses, investment and the cost of capital. A typical price is based on:

Concept	Amount	% of Equilibrium CAD Premium
Risk premium	140 €	70%
Internal expense & UCE	20 €	10%
Acquisition expenses	4 €	2%
Commission	30 €	15%
Investment income	-6 €	-3%
Capital cost (*)	12 €	6%
Total	200 €	

(*) Illustrative only. We assumed a solvency ratio of 50% and a return on capital of 12%.

As this example shows, actuaries put forth a lot of effort to calculate 70% of the final premium. A proper model should evaluate the impact of expenses, cost of capital and investment income. Costing is the part of the pricing process most widely used, but it is an inward-focused exercise and ignores what it is happening in the market. This cost-based approach does not take into consideration what a customer is willing to pay or what other carriers are charging for the same risk.

Competitive Market Analysis

Competitive Market Analysis (CMA) focuses on the comparison of rates offered by competitors. For personal lines, competitive pricing involves defining a portfolio of reference; that could be the existing book of business, a target portfolio or a portfolio that is representative of the insurance risks in a given market. How a reference portfolio is defined depends on the company's objectives. For example, when analyzing the overall market, we tend to use portfolios of one million risk profiles.

Because a company's competitive positioning can vary quickly in both hard and soft markets, it's important to frequently monitor competition. And rating structures are often not simple, so it is also important to assess the inner complexity of competitors' tariffs, which is the key to understanding their cross-subsidies. It is also very important to monitor new rating variables introduced by the market, trends in terms and conditions and changes in competitors' target markets. A company that prices its products low and that has a rating structure that implies a high degree of cross-subsidies will be in a very weak position in a soft market. A company that regularly monitors the market can use CMA to proactively manage the underwriting cycle through its various stages.

The data are analyzed using both univariate and multivariate analyses, which allow us to measure market competitive intensity, and a particular insurance company's position vis-à-vis its competitors, and then determine how to use this information to develop an informed pricing strategy.

Elasticity of demand

There are two aspects of the elasticity of demand that need to be considered: renewal rates (lapse analysis) and new business rates (conversion analysis). Several relevant variables can be used to identify customer behavior by risk segment, including the difference between an insurer's and competitors' premiums, price increases (valid only for renewals), the perceived value of the brand name, and of course, the characteristics (and segments) of customers who have a high propensity to switch insurance carriers.

New business intake could be analyzed by looking at conversion rates, which requires an appropriate conversion metric. One such metric is the proportion of policy quotes converted into sales. The conversion analysis examines the effects of customer behavior, marketing campaigns, and the sales process on the conversion metrics. Conversion pricing takes into consideration the premium levels offered by competitors, but also the elasticity of demand for products. Analyzing the key drivers behind product demand allows an insurer to adjust the price by segment to obtain higher conversion levels.

COMPETITIVE MARKET ANALYSIS

CMA has three relevant phases:

- **Data Gathering:** Different approaches are used in different countries to obtain competitors' rates depending on the data availability (e.g., rating books filed with regulators, quotes from agent/broker systems, industry association information and various ways of mystery shopping). These data have different degrees of reliability, and various adjustments may be needed before performing the analysis.
- **Univariate Analysis:** This methodology assesses the intensity of competition and company positioning, analyzing one variable at a time.
- **Multivariate Analysis:** This methodology analyzes all variables in combination to assess the intensity of competition and company positioning. Results help the company identify market niches and take rating actions to increase market share or minimize loss. This methodology is frequently performed through decision tree statistical techniques.

CMA techniques have, until now, been applied predominantly to motor insurance, mainly because in many countries this is the most significant line of business. However, the same concept can be applied to other personal lines products, such as health or household insurance.

CMA analysis is particularly effective under the following conditions:

- The product is treated as a commodity, with no guarantees or quality-of-service differences to differentiate competitor products.
- Premium rating structures are publicly available, and the actual premium charged by agents/brokers does not differ substantially from the rating book. (Discounts granted by agents can be as much as 30% for certain coverages.) It is possible to estimate agent discounts (through mystery shopping, for example) but the degree of accuracy is somewhat impaired.

- Competitor rating plans are based on known rating factors applied multiplicatively without caps or floors, and there aren't variables such as number of customers who have requested quotes, time of day, or quotes given before, during or after TV advertising.

UNIVARIATE ANALYSIS

Univariate analyses are used to present the results of the set of quotations in a simple and easy-to-interpret way.

Chart 1 represents each carrier's average premium, calculated on the overall portfolio. It is the first indication of the various companies' positioning compared to market averages and provides a quick indication of where competitors stand.

Chart 1

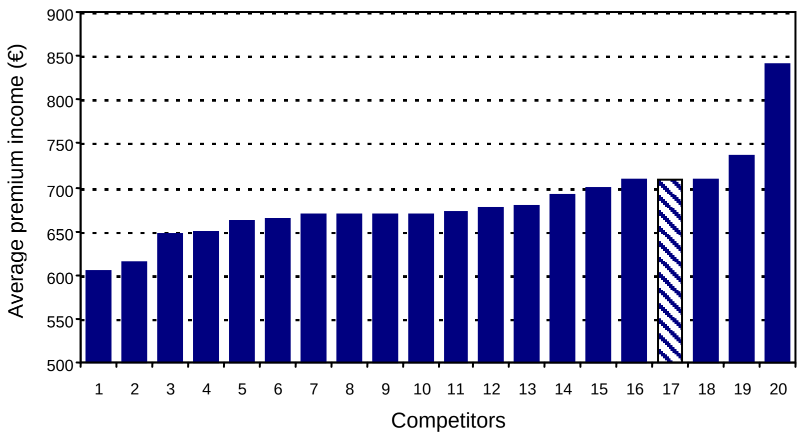
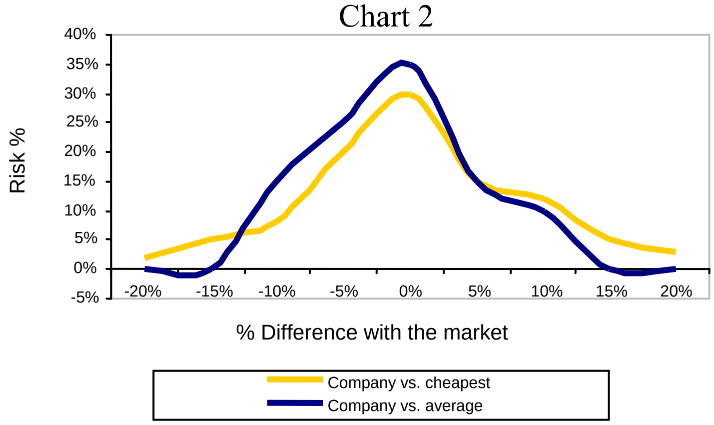


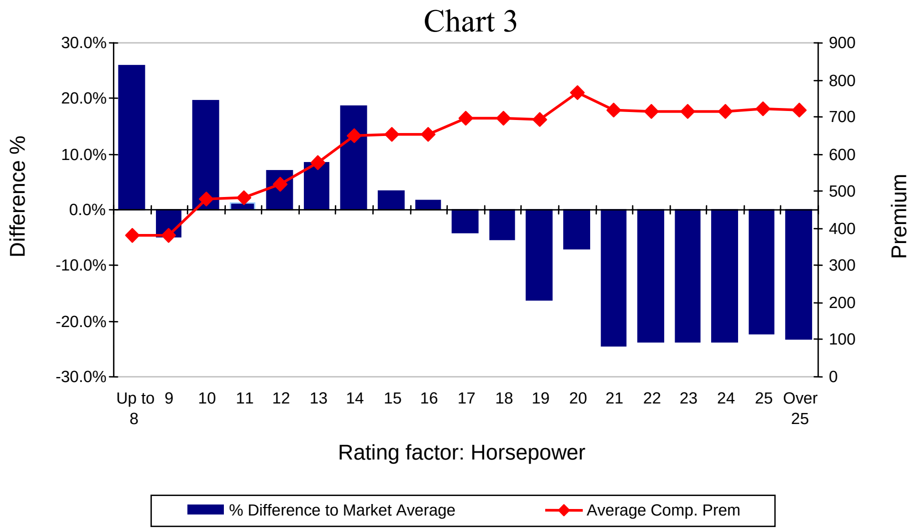
Chart 2: *Chart 1* provides a limited picture of the company's positioning. For example, an insurance company could be in line with the average market price overall, but its price distribution could be concentrated at the low or high end of the portfolio. It is therefore important to look at how differences in market price are distributed on the overall portfolio. *Chart 2* shows the percentage of cases in which the company is X% below/above average. In this example, for 15% of cases, the company is 10% below the average and, in 10% of cases, it is 10% above the average. The cross-hatch bar in *Chart 1* shows the same information relative to the lowest rate. In 3% of cases, the sample company is 20% more expensive than the cheapest company and, in 8% of cases it is 10% less expensive than the cheapest one.

Chart 2



By going to a deeper level of detail, we can analyze the company's competitive positioning by individual rating variables.

Chart 3 shows a comparison of premium based on horsepower. The x-axis shows horsepower levels, the left scale shows the percentage of difference between the company and the market for each horsepower group, and the right scale shows the average premium for each horsepower group.



The multivariate analysis also involves measuring the intensity of competition. We define intensity of competition as a measure of how hard insurance companies fight to win market share in particular segments. The closer prices are to each other in a segment, the greater the intensity of competition. Conversely, the more dispersed prices are in a given segment, the lower the competition. Intensity of competition is measured by the coefficient of variation (standard deviation/mean).

The multivariate analysis uses decision-tree techniques to assess an insurance company's price compared to the market average, in segments of different competitive intensity, for all rating variables at once. Results of the analysis can be used to develop different rating strategies to target specific market niches. In this paper, we use three combinations of intensity of competition (high, medium and low) with three levels of price difference compared to market average (in line, below and above market average) to create nine categorical levels of our dependent variable. The independent variable is all possible combinations of rating variables (i.e., customer segments in the reference portfolio under analysis).

Below is a brief overview of decision trees.

TREE-BASED METHODS

Description

The tree-based methods, or decision trees, are the most commonly used data mining techniques[1] in insurance and a number of other fields. They are particularly useful as an exploratory technique to discover relationship and structure in data, and they are appealing because they provide rules in human language, in contrast to regressions, which provide parameter estimates. For CMA, these rules help identify customers who fall in a particular category, in order to take the appropriate pricing action. In some contexts, the prediction is the only outcome that matters; the ability of the tree-based method to generate understandable rules that explain the reason for a situation or decision is important in the CMA context.

Tree-based methods are particularly suited for analyzing data with the following characteristics:

1. The partitioning of the relevant variables is disjunctive; therefore, an observation cannot belong to two categories at the same time.
2. The independent variables can be either categorical (e.g., ZIP code) or continuous (e.g., age).
3. The nominal response can have more than two levels (e.g., nine competitiveness-price clusters).
4. There is nonadditive behavior between the predictors and the response.
5. There are complex interactions between the predictors and the response.

Tree-based methods are suitable for both regression and classification problems, including:

- Regression trees where the dependent variable is continuous and the predictors are continuous or categorical. The output of a regression tree is in the following form:

    If driver age <= 25 and type of vehicle = small car, then

    Price difference to market average = -10%.

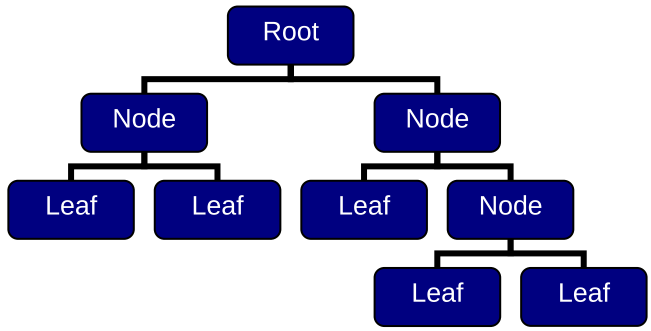
- For classification trees where the dependent variable is categorical and predictors are continuous or categorical, we use classification trees for CMA. There are a number of methods for analyzing these problems, including Logit regression, Log-linear analysis of multi-way frequency tables, C4.5-5.0, CHAID or CART. The output of a classification tree has the following form:

    If driver age <= 25 and type of vehicle = small car, then

    Price by competitiveness cluster = Low/High.

The CMA example that follows in the next section is a classification tree. The process to create a tree is called recursive partitioning, which is a top-down algorithm. A tree is represented by a flow chart *(Chart 4)* with a root node at the top, intermediate nodes and leaf nodes at the bottom. Links connect the nodes, and leaf nodes are the decision points. The decision trees are constructed by dividing the population into different nodes repeatedly until the stopping criteria are reached.

Chart 4



At each node, the aim is to identify the relevant predictive variable and its sets of values in order to assign an individual to a particular subgroup. The tree begins with one heterogeneous class and splits it into more homogeneous groups. Every predictive variable is considered for a split, and every combination of predictive values is considered for splitting a node. The ending splitting rule is obtained by comparing the result of the splitting measures of all possible splits.

The classic algorithms for a decision tree are:

CART (classification and regression trees) popularized by Breiman et al., 1984

CHAID (Chi-square automatic interaction detector) see Kass, 1980

C4.5/C5.0 (based on iterative dichotomiser 3, ID3) by Quinlan, 1993

Key elements of decision trees

The key elements of decision trees are:

- Splitting Search
- Number of Nodes.

Decision trees are constructed by splitting data into a number of nodes. These splits involve the selection of a partitioning (or splitting) variable whose values are used in designing the intermediate nodes. This process can be performed for continuous, nominal or ordinal predictors. Decision trees can be performed by binary recursive partitioning whereby one node is split only into two nodes. Binary splits are used frequently because they allow for an exhaustive search of possible splits, which are computationally more efficient than multi-way splits. For example, a variable with nine nominal levels would require evaluation of the splitting criteria 255 times, but the same variable in a four-way split would require 7,770 evaluations. The CART algorithm only allows for binary splitting, but the CHAID and C4.5/C5.0 allow for multivariate splitting.

- Splitting Criteria

The three most commonly used criteria for splitting a node when the response is nominal are "Gini," entropy and "chi-square". All criteria are based on the notion of impurity reduction. For continuous response, the sample variance can be used as a measure of impurity. The F test can also be used for regression trees. Classification trees are generally used for CMA. For this reason, criteria for continuous response, or regression tree, are outside the scope of this paper.

Gini (CART)

The Gini index is the measure of impurity of a node, and it is defined as:

$$g(t) = \sum_{j \neq i} p(j/t)\, p(i/t)$$

where j and i are categories of the response variable, and p(j/t) is the probability of category j at the node t.

If all cases in node t are of the same category (i.e., p(j/t)=1 and p(i/t)=0 for all i≠j), then the node is pure. The Gini index for a pure node is zero. As the number of evenly distributed categories increases (i.e., p(j/t)=1/n and p(i/t)=1/n for all i≠j, where n is the number of categories), then the node is impure and perfectly diverse. The Gini index approaches one when the number of categories increases for an impure and perfectly diverse node.

The tree searches for the split "s" that maximizes the formula below. The formula below is for binary split search. p_L is the proportion of cases of t sent to the left node; and p_R is the proportion sent to the right node. If the left and right nodes are pure, then $g(t_L)=g(t_R)=0$ and $\Phi(s,t)=g(t)$. That case is the best possible split. The formula searches for the split that reduces the impurity of a node.

$$\Phi(s,t) = g(t) - p_L g(t_L) - p_R g(t_R)$$

The tree will stop splitting when we meet one of the stopping rules described later, or when the node is pure.

Entropy (C5.0)

Let n_b be the number of instances in branch b

Let n_{bc} be the number of instances in branch b of class c, therefore n_{bc} is less or equal to n_b

If $n_{bc}/n_b = 1$ then all the instances are from class c, and we have a homogeneous node. If the ratio is zero, there are no instances from class c, and we have a heterogeneous node.

$$Entropy = \sum_c -\left(\frac{n_{bc}}{n_b}\right)\log_2\left(\frac{n_{bc}}{n_b}\right)$$

In this case, when n_{bc}/n_b gets close to zero (i.e., the node has only a few instances of class c) then the $\log_2(n_{bc}/n_b)$ tends to be a large negative number. Similarly, if n_{bc}/n_b gets close to one (i.e., the node has most of the instances of class c) then the $\log(n_{bc}/n_b)$ approaches zero. Hence, we see that when a class is nearly, or completely, representing all the instances, the entropy tends also to zero.

Entropy tends to look for splits where as many levels as possible are divided correctly. As a result, entropy puts more emphasis on categorizing the most number of classes correctly, while the Gini index favors isolating the class with the largest population in one node.

X2(CHAID)

A split can be represented by a contingency table. The rows represent the chaid nodes, and the columns represent the classes. The chi-squared test can be used to estimate the significance of a split. Here is an example of a contingency table for the observed, and expected, competitiveness for a potential split.

Observed Competitiveness	Driver age <=25	Driver age > 25
Low	293	71
Medium	363	1
High	42	294

Expected Competitiveness	Driver age <=25	Driver age > 25
Low	239	125
Medium	239	125
High	225	116

The Pearson chi-squared test is defined by

$$X_v^2 = \sum \frac{(Observed - Expected)^2}{Expected}$$

where v is the number of degrees of freedom defined by (number of rows - 1) (number of columns - 1). The number of rows is the number of target classes (i.e., low, medium, high), and the number of columns is the number of categories from a split (two in the case of a binary tree).

For the example above X^2 is 644.

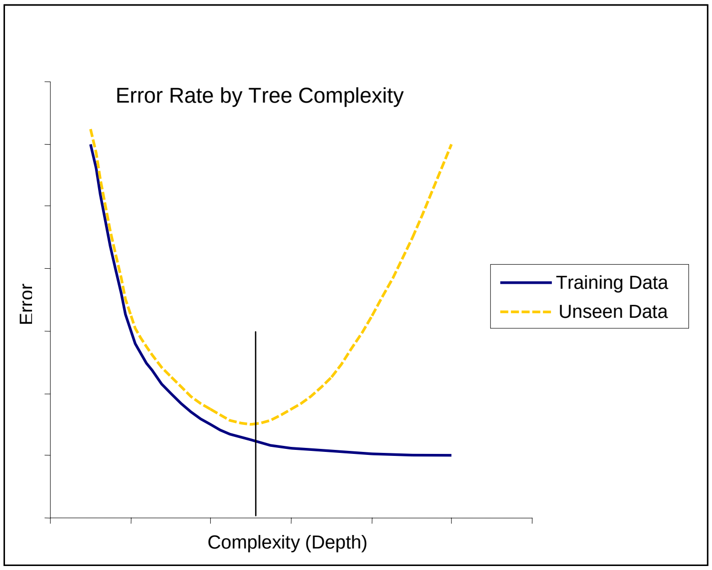
$(Observed - Expected)^2/Expected$	Driver age <=25	Driver age > 25
Low	12	23
Medium	64	123
High	149	273

The Gini index and the entropy tend to increase as the number of branches (e.g., two for binary tree) increase, although they do not have an adjustment similar to the number of degrees of freedom. Therefore, the Gini and entropy criteria are not adequate for multivariate splits evaluation.

■ The Stopping Measures

The stopping rules determine when the recursive splitting process stops. If not stopped, the tree algorithm will ultimately divide the tree to achieve a single observation in each node. This means that the algorithm will extract all the information from the data, including the random noise. To avoid this situation, stopping measures are implemented in the tree algorithm. The decision tree needs to be optimized for a more general population than the pattern found in the data. *Chart 5* shows the trade-off between the tree complexity, defined by the tree depth, and the error on unseen data (i.e., out of training sample).

Chart 5



The rules of stopping could be predetermined with the following basis:

Minimum: Control the number of cases in a node.

Depth of the tree: Determining a priori the maximum number of splits that we will accept.

■ The Pruning Process

Pruning can be performed with two different approaches:

1. Pre-pruning. Allow the tree to grow to just the right size, where the right size is determined by the user, based on the knowledge from previous research or analysis.
2. Post-pruning. Typically we can use four diagnostics to assess the goodness of fit:
   — *Test sample cross-validation.* The tree is computed from a training sample, and its predictive accuracy is tested by applying it to predict the class membership in the test sample.
   — *Cross validation.* The data set is randomly split into N sections, and the tree computed n times, each time excluding a section as test sample, then proceeding as in the previous test, using the average of the n test.

— *Minimal cost-complexity cross validation pruning.* For a splitting criteria S, a penalty parameter is introduced to take into account the model complexity of a subtree T.

$S_p(T) = S(T) + p \text{ size } (T)$

— *Select a ratio of cases that are correctly assigned to each node.*

This analysis will help identify those terminal nodes where the misclassification rates are below a "hurdle" level. These misclassification levels (or error rates) are then standardized by the number of nodes in the tree to adjust the error rate.

The objective of pruning is to produce a simpler tree, following the Occam's Razor principle, which states that "the explanation of any phenomenon should make as few assumptions as possible, eliminating, or 'shaving off,' those that make no difference in the observable predictions of the explanatory hypothesis or theory. When given two equally valid explanations for a phenomenon, one should embrace the less complicated formulation."[2]

Advantages and disadvantages of decision trees

The advantages of decision trees are that they:

- handle continuous, discrete, categorical variables
- have the ability to deal with incomplete or missing data
- have no need to make assumptions about the distribution of the variables
- are invariant to extreme values, collinearity or heteroschedasticity

The disadvantages of the decision trees are that they:

- are error-prone with too many classes
- have difficulty representing a linear relationship.

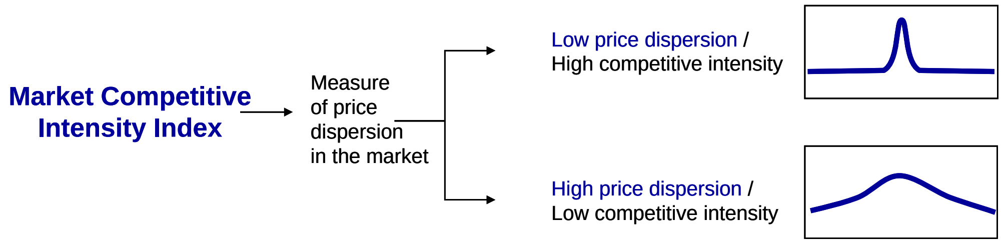
CLASSIFICATION TREES APPLIED TO CMA

In this section, we provide an overview of our methodology to help insurance companies better define their pricing strategy on the basis of market intensity of competition and difference of price to market average. This methodology poses no particular limit (other than current computer and software capability) to the complexity of the rating structures and the size of the dataset to be analyzed.

Intensity of Competition

Intensity of competition measures the price dispersion for a given group of risk. Higher dispersion means lower intensity of competition and lower dispersion means high intensity of competition. It is measured by the coefficient of variation and is simplified by the chart below *(Chart 6)* and is predefined into three different categories: high, medium and low intensity of competition.

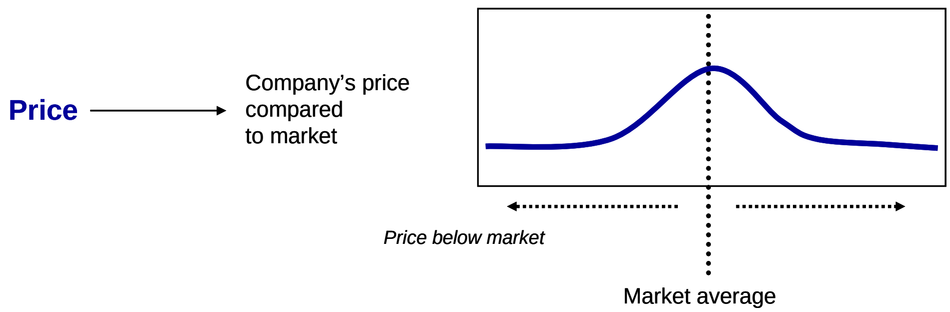
Chart 6



Price difference to market average

The price index compares the company's rate for a set of risks against the market price for this set. The price can be "below market average," as shown in *Chart 7*, "in line with" or "above the market average."

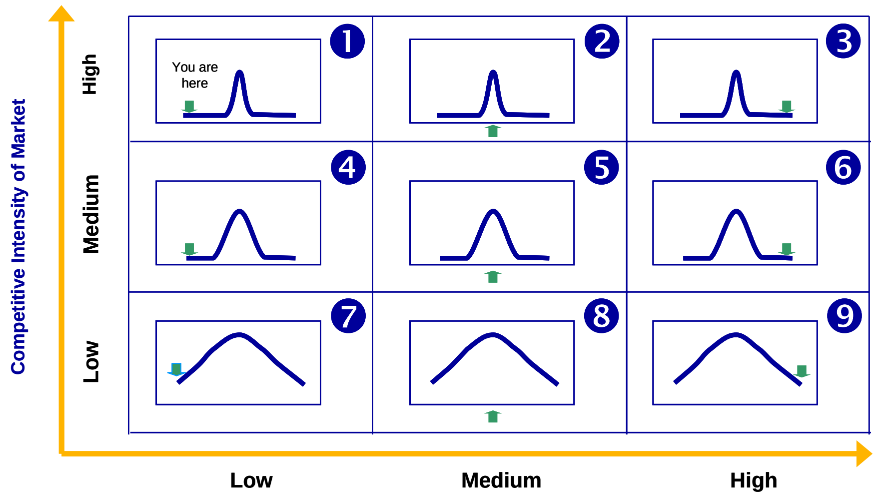
Chart 7



Price Competitiveness Clusters

Each customer in the database is assigned to one of the nine combinations of price/competition. Decision trees are then applied to determine the most explanatory rating variables to an insurance company market positioning in respect to intensity of competition. Each of the clusters identified suggests potential pricing strategies and rating actions. In each of the clusters resulting from the model, we will be able to assess not only our positioning against the market average, but also determine who the main competitors are. Our main interest is to analyze clusters in which our price is below the market average, to find niches where we would be able to increase price with minimal risk of losing market share. For risks falling into cluster 1 (*Chart 8*), we have greater opportunities to increase price because we are particularly low-priced and the high intensity of competition suggests that prices are very close together. Therefore, existing customers who shop around will find that prices are similar to what they are already paying, making them less likely to switch insurers.

Chart 8



In clusters where price is still below the market but where the intensity of competition is decreasing (clusters 4 and 7 of the matrix), we still have the opportunity to increase price, but by a lower magnitude. This is because the higher price dispersion increases the probability that customers who shop around will find much cheaper quotes and decide to switch. By properly identifying risks falling in clusters 1, 4 and 7, we can establish price increases that minimize the risk of losing market share.

From a strict methodological point of view, this analysis does not necessarily require an assessment of the level of profitability, since we are simply increasing rates. However, other considerations must be taken into account, particularly when increasing prices in already profitable segments. Depending on the magnitude of the price increase, we can define a strategy to increase revenues and profit while minimizing the risk of losing market share. However, we will need to perform an elasticity-of-demand analysis to quantify how much we will lose. Then, using this knowledge, we can optimize the rating structure.

Where our rates are particularly high (3, 6 and 9), we can gain market share without impairing underwriting results by lowering our price, depending on competitive intensity and taking profitability into account. Again, we could use an elasticity-of-demand analysis to maximize new customers' conversion rates.

We provide an example of the output of our analysis in the charts below. *Chart 9* summarizes the difference to market averages (in circles) and the percentage of risk profiles falling in those clusters (rectangles). The dotted line suggests the ideal positioning of one company based on the intensity of competitive market dynamics. The number of risk profiles falling into each cluster depends on the overall market positioning of the company. (For example, a very expensive company will have most of its risks on the right side of the matrix.) But it will also depend on the implied level of complexity of the company rating structure and its ability to target specific segments. In the case illustrated in *Chart 9,* it could be that very few risk profiles will be in the middle of the matrix and that the distribution of the price differences will be concentrated in the tails (very expensive/very cheap).

Chart 9

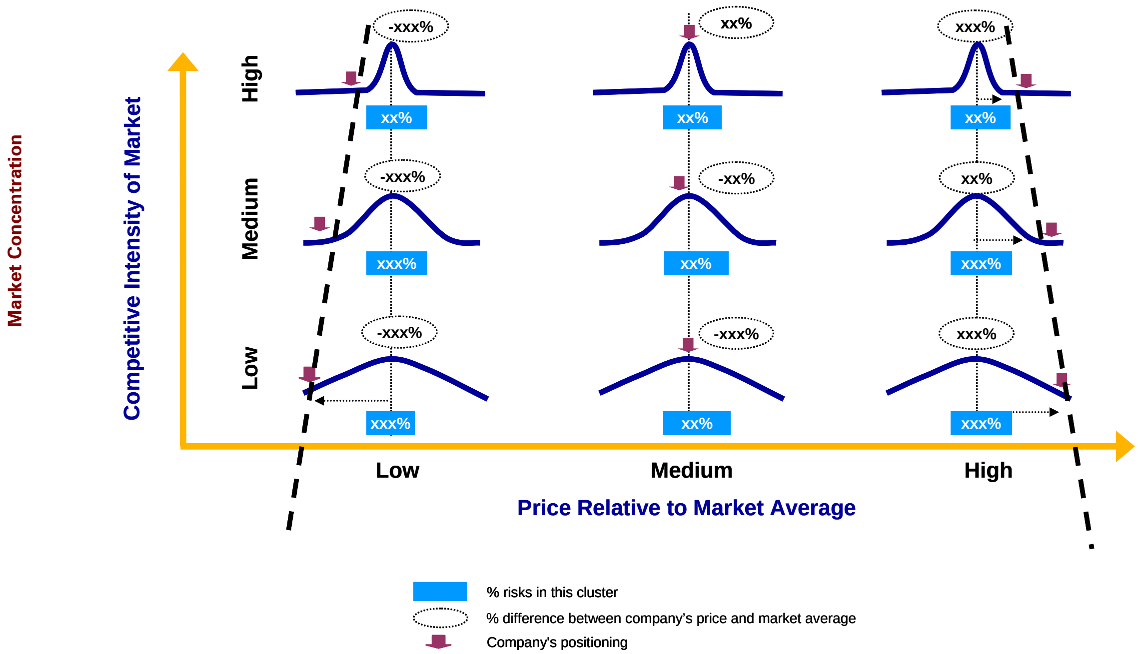
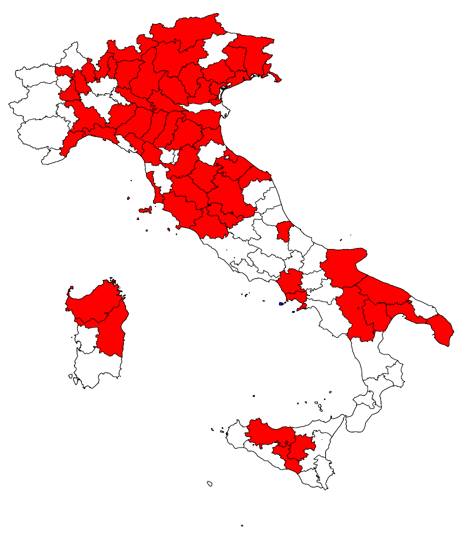


Chart 10 is an example of the output of the analysis that describes the risk segments in one of the clusters.

Chart 10

Rating Variable	Values
Age	>40 years
Gender/ Marital status	Female/ married
Miles to work	15 miles
Model years	New car
Installments	2 yearly installments

Shaded areas on map represent geographical areas that fall in this node.



The chart above shows one of the nodes resulting from the decision tree algorithm. It is representative of risk profiles falling into cluster 1 of the above table (high intensity of competition/low price) where it is possible to increase price while minimizing the risk of losing market share. In this example, the risk profiles are customers living in ZIP Codes identified in shaded areas, females older than 40, who drive 15 miles to work in a new car and pay their premium in two yearly installments.

CONCLUSION

Pricing Optimization Management should balance cost, competitiveness and elasticity of demand, as well as several other elements, to achieve the company's long-term profitability and growth objectives. To perform this process, it is necessary to use an integrated analytic pricing methodology that takes into consideration all aspects of the rating process.

CMA and the integrated approach to pricing described in this paper have been adopted by insurance companies to grow market share without compromising profitability. The motor market is increasingly competitive everywhere, and advanced statistical techniques are needed to successfully manage a motor portfolio. In the near future, we will see more and more companies adopting the approach described in this paper. First-movers will, however, have a considerable advantage.

ENDNOTES

R. Lewis *An Introduction to CART Analysis.* Presented at the 2000 Annual Meeting of the Society for Academic Emergency Medicine in San Francisco, U.S.A.

J. Salinas *Patrones de Morosidad para un producto crediticio usando la técnica CART.* Revista de la Facultad de Ingenieria Industrial, Vol (8).

K. Bauer *The power of Metrics: Predictive Analytics: Branching our with decision trees.* April 2006 *DMReview.com.*

R. Phillips *Pricing and Revenue Optimization.*

Breiman, Friedman, Olshen& Stone 1984 *Shape Quantization and Recognition with Randomized Trees.*

Quinlan, J. R. C4.5: *Programs for Machine Learning.* Morgan Kaufmann Publishers, 1993.

M.J.A. Berry, G. Linoff, *Data Mining Techniques: For Marketing, Sales, and Customer Support*. Wiley, 1997.

Hastie T., Tibshirani R., Friedman J. *The Elements of Statistical Learning,* Springer, 2001.

W.J.E. Potts, *Decision Tree Modeling Course Notes,* SAS Institute, 2000.

Kass, G.V. (1980) *An Exploratory Technique for Investigating Large Quantities of Categorical Data,* Applied Statistics, 29 119-127.

Hadidi, Nasser (2003) *Classification Ratemaking Using Decision Trees,* CAS Forum, 258-283.

S-PLUS 7 *Guide to Statistics,* Volume 2, Insightful Corporation, 2005.

Ruoming J., Gagan A *Efficient Decision Tree Construction on Streaming Data.*

---

[1] page 7, *http://www.kdnuggets.com/polls/2006/data_mining_methods.htm*
[2] Page 12, from Wikipedia (*http://en.wikipedia.org/wiki/Occams_razor*)