MIE1612 Project:
Optimal Pricing for Insurance Products Using Multistage Stochastic Programming

Paul Xie

December 2021

Abstract

Figuring out the best prices for different types of products has received increasing amount of attention in the financial sector recently. Traditional stochastic programming models with recourse usually assume the uncertainty to be exogenous. However, due to the complexity of pricing problems, customer demands is always highly related to the prices of products. In this project, I proposed an optimal pricing model for insurance products so that the overall profit for the insurance company is maximized. I used Monte Carlo simulation to generate customer demands and accidents for a sample size of 200 individuals. A multi-stage stochastic programming model is studied and solved to optimality using the commercial solver Gurobi. Finally, the experimental results and discussion about future improvements are presented.

1 Introduction

Today's world is more competitive than ever before. For this reason, revenue management and pricing policies are the most effective tools that companies can use to manipulate market demand and maximize potential benefits. Setting the price too high will result in less customers, while setting the price too low will not the selling profitable for the company [1]. For insurance companies, we also need to take into account the liability rates, as profit is largely correlated to the number of claims made by the customers. For instance, auto insurance product A is relatively cheaper but it does not cover minor accidents, while product B is more expensive and covers all accidents. In the short term, selling more product B might seem more profitable, but it might potentially result in larger loss to the company due to claims. Thus, a multi-stage stochastic programming model shall be used.

Most practical decision programming involving uncertainty at some level, and the concept of stochastic programming was introduced by Dantzig and Beale [2] in 1955. However, the realization of their approach requires the uncertainty to be independent from decision variables, and is represented by a finite set of scenarios. Recently, stochastic programs with endogenous uncertainties or decision dependent uncertainties have received increased attentions, but to the best of my knowledge, there has been little to no study where the relation between random variables and decisions is modeled as a continuous function. Because of this, we need to create a finite set of pricing options and simulate the random variables using Monte Carlo Methods.

The main contribution of this project is to provide a new formulation for endogenous stochastic programming models that is able to find the optimal prices for multiple insurance products for the insurer, while taking into account premium collections. The structure of this report is as following: Section 2 will formally introduce the problem I was solving, Section 3 will introduce the mathematical formulation of the stochastic programming model, Section 4 will present my experimental results, and finally Section 5 will discuss how this model can be generalized and where it can be improved.

2 Problem Definition

In the auto insurance industry, customers' willingness to purchase a certain insurance product is largely dependent on the price and coverage. As such, an insurance company (called insurer for the remainder of this report) can offer multiple auto insurance products at different price points. Customers' price sensitivity are, intuitively, different toward each product. For instance, if a product has better coverage, then the customers are willing to spend some extra money for the better service. Sometimes, a customer might be interested in more than one products provided by the insurer, which means the insurer can decide which product should be offered to the client based on their likelihood of getting into accidents. More expensive products means the insurer can make more money from collecting regular monthly payments, but it also means they will lose more once the client run into accidents.

In this project, I assumed the insurer offers three different types of car insurance products A, B and C, with A being the cheapest and C the most premium. There are two sets of decision variables: the prices to set for each of the three products, and the products to offer for the consumers. There are also two sets of random variables: customers' desire to purchase each of the products, and the likelihood of them getting into accidents. The problem aims to look for the optimal decision variables such that the expected profit is maximized for the insurer.

3 Method

The biggest challenges of modelling a pricing model is the endogenous random variables. The decisions we make will have a substantial consequence on the uncertainty we will face later. There are two types of decision-dependent uncertainties in stochastic programming [3], and in this case, every different decision we make will result in a different probability distribution for the random variable (e.g. demand). Various methods of modelling such uncertainties has been studied [3] and I decided to use Monte Carlo simulation for all the random variables, and use a multi-stage stochastic programming model to solve for the problem.

3.1 Mathematical Definition

Let $J = \{A, B, C\}$ be the set of all car insurance products available from the insurer. I is the set of all potential customers on the market, and P is a finite set of pricing options. As every different pricing strategy will result in different probability distribution for customer demands, let's use Pc to denote the set of all possible pricing combinations for products A, B and C. For instance, if we have 10 pricing options (e.g. $|P| = 10$), then $Pc = \{(p1, p2, p3), (p1, p2, p4), (p1, p2, p5), \cdots, (p8, p9, p10)\}$. Let S_1 and S_2 be finite sets of random scenarios for the first and second stage respectively. S_1 is related to the fluctuation of customer demands Customer demands, it could include scenarios such as tighter governmental requirements over car insurance, or more reported car thefts, which will likely result in a change in demands. S_2 denotes the random scenarios of weathers. This is directly correlated to the number of accidents and the number of claims made by customers. Good weather will likely witness less car accidents while bad weather will surely increase accidents occurrence rate. Demand is denoted by $d_{ij}^{p_c,s_1}$, which is a binary variable. $d_{ij}^{p_c,s_1} = 1$ indicates customer i is willing to buy product j given the pricing strategy p_c and first stage random scenario s_1. Let $A = \{S, M, L\}$ denote the set of all accidents. Accidents are classified into three categories: Small, Medium and Large. Each type of accidents requires a different amount of payment from the insurer, and the actual amount of premium the customers collect also depends on the type of product they purchase. Use C_{ja} to denote the cost for the insurer if a customer buys product $j \in J$ and gets into accident type $a \in A$. Finally, $\alpha_{i,a}^{s_2}$ denotes the number of time person $i \in I$ gets into accident type $a \in A$ under weather condition $s_2 \in S_2$.

The variables used in this model are:

- $x_{jp} \in \{0, 1\}$: equals 1 if set the price of product $j \in J$ to price $p \in P$
- $r_j \in P$: the actual sell price for product $j \in J$
- $y_{ij}^{p_c,s_1} \in 0, 1$: equals 1 if assign product $j \in J$ to customer $i \in I$, with pricing policy $p_c \in Pc$, under random scenario $s_1 \in S_1$

Overall information flow in this problem can be visualized by Figure 1. Variables x and r are first stage decision variables, y is a second stage decision variable and d, α are random variables which will be further discussed in Section 3.2. Even though there are random variables in two stages, there is no third stage decision variable in this model.

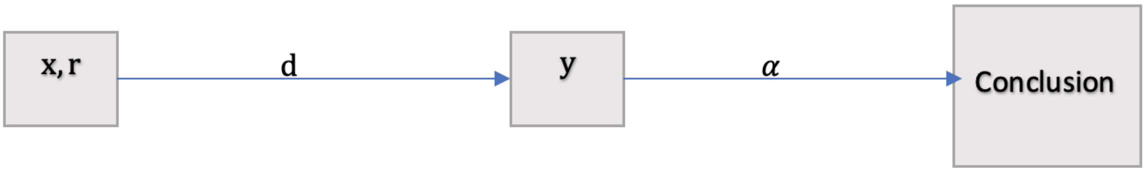

Figure 1: Information flow in the problem. The variables in the boxes are decision variables, the ones on the arcs are random variables.

The objective function in this problem is:

$$max \quad \mathbb{E}_{s_1}[\sum_{p_c}\sum_{i}\sum_{j} y_{ij}^{p_c,s_1} \cdot r_j - \mathbb{E}_{s_2}(\sum_{a}\sum_{p_c}\sum_{i}\sum_{j} \alpha_{ia}^{s_2} \cdot y_{ij}^{p_c,s_1} \cdot C_{ja})] \tag{1}$$

The first portion of equation 1 corresponds to the revenue of the insurer from selling the products, and the second part of the objective function corresponds to the amount loss by the insurer from customer claims. We will be maximizing equation 1 subject to the following constraints:

$$y_{ij}^{p_c,s_1} \leq d_{ij}^{p_c,s_1}, \quad \forall i \in I, j \in J, s_1 \in S_1, p_c \in Pc \tag{2}$$

$$\sum_{j} y_{ij}^{p_c,s_1} \leq 1, \quad \forall i \in I, s_1 \in S_1, p_c \in Pc \tag{3}$$

$$r_A \leq r_B \leq r_C \tag{4}$$

$$\sum_{p} x_{jp} = 1, \quad \forall j \in J \tag{5}$$

$$r_j = p \cdot x_{jp}, \quad \forall p \in P, j \in J \tag{6}$$

$$3 \cdot y_{ij}^{p_c,s_1} \leq x_{A,p_A} + x_{B,p_B} + x_{C,p_C}, \quad \forall i \in I, j \in J, p_c = (p_A, p_B, p_C) \in Pc, s_1 \in S_1 \tag{7}$$

The first constraint (eq (2)) makes sure that the insurer may only assign product j to person i if the consumer wants it. In other words, the insurer cannot force the consumers to purchase a product they do not want. Constraint (3) restrains the insurer to sell at most one product per customer. Constraint (4) guarantees product A is the cheapest and C is the most expensive. Constraint (5) tells that the insurer must set one price for each of the three insurance products. Constraint (6) is the linking constraint between x and p. And constraint (7) mean the insurer can assign product j to customer i under a certain pricing strategy, if and only if that strategy is taken.

Note that this formulation is a quadratic programming since both y and r are decision variables. However, all the constraints are linear with the introduction of r.

3.2 Monte Carlo Simulation for Random Variables

Customer demands based on prices and accidents should be simulated [4]. One big advantage of having a finite set of pricing options is that there are finitely many possible outcomes for demands, thus can be simulated. A continuous set of pricing options will result in the demand to be a continuous function of initial decision variables, which will have infinite number of possible outcomes thus cannot be simulated. The relation between customer demand and price has been proposed in [5], and the overall demand in the population can be formulated as:

$$D_j^{s_1} = F(r_j, s_1) = N - b_j \cdot r_j + \epsilon_{s_1} \tag{8}$$

In equation (8), $D_j^{s_1}$ is the total number of demand from the entire market for product j under random scenario s_1. On the right hand side of equation (8), N corresponds to the market size, b_j is the price sensitivity towards product j, and ϵ_{s_1} is a random variable under scenario s_1. Therefore, $\frac{D_j^{s_1}}{N}$ will give us the proportion of customer that would consider buying product j.

For the simplicity of this project (and also time constraint), I assume $N = 200$, and chose reasonable values for b_j and ϵ_{s_1}. I have also assumed that all 200 people in the market are completely the same, which means there is no difference in terms of income level, preference or driving habits (otherwise it will make the simulation way too complicated, which shouldn't be the focus of this project). Once we know the proportion of customers who wants to buy certain products, we can use binomial distribution to sample customer demand since people's desire for a certain product has binary outcome. Figure 2 is the result of the sampling if the insurer decides to use pricing strategy (p_2, p_8, p_{11}), and the first stage random variable is in the state s_{13}.

	A	B	C
0	1	0	1
1	1	0	0
2	1	1	1
3	1	0	0
4	0	0	1
...	...	...	...
195	0	0	1
196	1	1	1
197	1	0	0
198	1	0	1
199	1	0	0

Figure 2: Simulated demand with pricing strategy (p_2, p_8, p_{11}), under random scenario s_{13}.

Intuitively, people tend to be more sensitive towards the product with only "base coverage", and are willing to pay some extra for the more premium ones. As we can see from Figure 2, a person can be interested in purchasing more than one insurance products, but the insurer may only sell at most one to each customer.

As for the second stage random variable, it is fortunately decision-independent (exogenous), and the likelihood of accident is only decided by weather ($s_2 \in S_2$). Again, I assumed no prior knowledge of customers' driving habits and claim history, which means all of the 200 people are equal likely to get into any type of accidents. Figure 3 shows the simulated accidents for each customer under "average" weather. Note that the customers might get into multiple accidents of the same type during the year.

	small	medium	large
0	2	1	0
1	1	1	0
2	1	0	0
3	2	1	0
4	2	0	0
...	...	...	...
195	1	0	0
196	1	0	0
197	0	1	0
198	1	1	0
199	2	0	0

Figure 3: Simulated accidents under "Average" weather.

Larger accidents require larger amount of premium collected by the customers, but the amount of money the insurer actually needs to pay also depends on the type of insurance product the customers purchase.

4 Experimental Results

For the computational experiment, I set $N = 200$, $P = \{50, 100, 150, 200, \cdots, 500, 550, 600\}$. I assumed there are four possible scenarios in S_1, with probability equal to 25% each. The second stage exogenous random variable S_2 has 20%, 60%, 20% chances of being "bad", "average", "good" weather respectively. The costs of accidents based on the insurance plan C_{ja} is summarized in Table 1. Price sensitivities b_j for A, B and C are set to 0.6, 0.4, 0.2 respectively.

	Small	Medium	Large
A	100	1000	3000
B	300	1500	4000
C	500	2500	5000

Table 1: Costs for different types of accidents based on insurance products purchased.

The first stage randomness ϵ is equal to $\{4, 100, 20, 30\}$ for the four random scenarios in stage 1. Indeed, there should be two random variables in stage one, one exogenous the other endogenous, since the equation (8) has stated that the demand is a function of both price and randomness (ϵ). The second stage random variables S_2 can be summarized in Table 2. For all simulations on the uncertainty, the random seed is set to 1612 to guarantee consistent results.

	Small	Medium	Large
s_{21} ("bad")	0.45	0.3	0.1
s_{22} ("average")	0.4	0.2	0.05
s_{23} ("good")	0.35	0.15	0.02

Table 2: Chances of getting into each type of accident under different weather conditions.

The model described in Section 3.1 was implemented the commercial solver Gurobi, and it is solved to optimality. The machine I used to solve the model has the hardware specs of 2.6 GHz 6-Core Intel i7, with 16 GB of RAM DDR4.

The optimal solution found for the first stage decision variable r is:

$$r^* = \{300, 500, 550\} \tag{9}$$

indicating the best way to set price is to sell A at 300, B at 500 and C at 550. After the revelation of first stage random variables (demand), we would assign the second stage decision variables y to the following values:

- scenario 1:

  Sell product A to customer 31;
  Sell product A to customer 129;
  Sell product A to customer 140;

- scenario 2:

  Sell product A to customer 31;
  Sell product B to customer 36;
  Sell product B to customer 124;
  Sell product A to customer 129;
  Sell product A to customer 140;
  Sell product B to customer 143;
  Sell product A to customer 147;
  Sell product A to customer 155;
  Sell product A to customer 159;
  Sell product A to customer 170;
  Sell product A to customer 179;
  Sell product B to customer 188;

- scenario 3:

  Sell product B to customer 124;
  Sell product A to customer 140;
  Sell product C to customer 155;
  Sell product B to customer 188;
  Sell product A to customer 189;

- scenario 4:

  Sell product A to customer 31;
  Sell product B to customer 36;
  Sell product A to customer 155;
  Sell product B to customer 159;
  Sell product A to customer 170;
  Sell product A to customer 189;

The optimal objective value is 767.5, which means the insurer will earn 767.5 amount of net revenue from car insurance products. More details about the experimental results, please refer to the complimentary Jupyter Notebook.

5 Conclusion

In short, the proposed model in this project is able to solve the optimal pricing problem for insurance products, given enough information about the relation between customer demand and price, as well as accident rates. However, the time it took to solve the problem to optimality is tremendously long. With the computational specs that I mentioned in Section 4, it took nearly 20 minutes for the solver to find optimality. This is primarily due to the fact that the set Pc is very large, since it contains all pricing strategies (combinations). If the insurer offers three distinct products, then $|Pc| = |P|^3$, where P is the set of all pricing options. In this project, I assumed 12 possible pricing options, which means $|Pc| = 1728$. Because of this, there are not only a large number of constraints (e.g. $|y| = 1728 \times 200 \times 3 \times 4 = 4147200$), but also a large number of constraints (equations (2)(3)(7)). In real-world situations, there is likely to be far more than 200 people in the market, and we might potentially have more than just 12 pricing options to choose from. In order to have this formulation applicable for more realistic problems, additional algorithms such as Bender's decomposition [6] should be used to reduce solution time.

In terms of the Monte Carlo simulation, we can further cluster the customers into different income level, different age groups and different preferences to make it more realistic. People with higher income usually tend to be less price sensitive compare to people with lower income, and are more willing to spend on more expensive products. Additionally, it is very helpful for the insurers to check customers' driving experience, driving habits and past claim history to have a more accurate expectation about their chances of getting into accidents. The relation between pricing and customer demand deserves further study, and it is likely to involve an investment into incomplete markets [7]. Statistical surveys and machine learning approaches can also help to determine the relation between price and demands.

Future study in this field should focus on techniques that will improve the solution time of the proposed model, or obtaining more realistic data (or simulation) which makes the problem more realistic.

References

[1] PhD Richard V. Eastin and CFA Gary L. Arbogast. Demand and supply analysis: Introduction. *CFA Institute*, 2011.

[2] Julia Higle. *Stochastic Programming: Optimization When Uncertainty Matters.* 09 2005.

[3] Lars Hellemo, Paul I. Barton, and Asgeir Tomasgard. Decision-dependent probabilities in stochastic programs with recourse. *Computational Management Science*, 15(3):369–395, Oct 2018.

[4] Reuven Y. Rubinstein and Dirk P. Kroese. *Simulation and the Monte Carlo Method, Third Edition.* Nov 2016.

[5] S.K. Chaharsooghi, M. Honarvar, and M. Modarres. A multi-stage stochastic programming model for dynamic pricing and lead time decisions in multi-class make-to-order firm. *Scientia Iranica*, 18(3):711–721, 2011.

[6] Ragheb Rahmaniani, Teodor Gabriel Crainic, Michel Gendreau, and Walter Rei. The benders decomposition algorithm: A literature review. *European Journal of Operational Research*, 259(3):801–817, 2017.

[7] Sergei Fedotov and Sergei Mikhailov. Option pricing for incomplete markets via stochastic optimization: transaction costs, adaptive control and forecast. *International Journal of Theoretical and Applied Finance*, 4:179–195, 02 2001.