Stockholms universitet

Commitment and discounts in a loyalty model

Martin Karvik

Masteruppsats i försäkringsmatematik
Master Thesis in Actuarial Mathematics

Masteruppsats 2016:1
Försäkringsmatematik
Juni 2016

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Master Thesis **2016:1**
http://www.math.su.se

Commitment and discounts in a loyalty model

Martin Karvik*

June 2016

Abstract

The loyalty of an insurance customer is usually defined from how long a customer decides to stay and renew the insurance company's services or products. There are several ways to develop such a model. By building a model for loyalty, we can explore the properties of the customers and how they act and react to different situations depending on their age, type of object insured, how long they have been insured and so on. Most companies use retention programs to make the customer invest in more products, but also to stay longer than they usually should have by making the customer commit to more products and services. To make this an attracting offer for the customer, it is sometimes combined with a discount. But what matters most for loyalty; the discount or the commitment (representing convenience, lock in effects, satisfaction, etc)? To do this we studied car insurance at Länsförsäkringar, a federation of 23 different regional companies sharing a common brand and regulating their own retention programs and discount levels. The aim of the paper is to build a full model describing the loyalty of non-life insurance customers, using all eligible and relevant factors, and then to separate the loyalty added from commitment into what is price related and what is not. Conditioning a discount on a commitment can be an effective way to increase the loyalty of a customer. If the expected time in years a customer without commitment will stay in Länsförsäkringar is two years, then the partially committed with a discount of 5% is expected to stay three years and the fully committed with 19% six years. A customer without commitment needs a discount of 23% to be as loyal as a fully committed customer with 15%.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: martin.karvik@gmail.com. Supervisor: Filip Lindskog.

Table of contents

1 Introduction

1.1 Objective

The aim of this paper is to explore customer loyalty from a retention program in an insurance company. By investigating the contribution from price and commitment into other products or services, we will try to find out how much they influence the loyalty separately and together. To do this, we need to build a model for the loyalty. Every step will be thoroughly motivated and explained. Finally, we will try to explain how the results can be interpreted in terms of other loyalty measures that are easy to understand and relate to. The product that we are modeling is car insurance.

1.2 What is loyalty?

Whether you are a restaurant owner, a producer of goods or a financial institute, you need to have a strategy to make your profitable customer return to buy your goods or keep using your services. We usually call a customer that returns often as a loyal customer. But loyalty has many meanings and needs to be defined to really understand the scope and purpose of this paper.

Let us start with how we do not define loyalty. A customer's feelings towards a certain brand or product or how often a customer recommends the product to friends, sometimes called ambassadorship, are really hard to measure and would require a survey on its own. Surveys as such are time consuming, difficult to get enough data points and with the risk of getting biased. No surveys have been conducted and we are not to define loyalty in that way. A customer that broadens or deepens the commitment in products or services from the same brand is also excluded as a definition of loyalty. As we are studying car insurance, a deepening of the commitment would mean to buy more cover for the car, or to insure multiple cars. That is more an indication of wealth or worry, than actual loyalty. Broadening the commitment into more different products within the same brand might be considered as loyal behavior, but will not be studied. The focus will be car insurance and loyal behavior will be measured for car insurance. Having a broad product flora, or commitment as we will call it from now on, will indirectly be considered though, since it has an effect on the loyalty of a car insurance customer.

A loyal customer stays with the same company for a long time, compared to a disloyal customer that only stays for a short period of time, perhaps looking for the cheapest offer among all competitors. To stay for a long period of time as a customer of car insurance, the customer needs to renew the contract yearly, since car insurance contracts for the private line lasts only a year. The more times a contract are renewed, the longer a customer will stay and the more loyal is the customer. Hence, loyalty refers to the duration of a customer in a given product.

1.3 Why is loyalty important?

Insurance companies operate in a more competitive environment now than they used to do in the past. Customers easily switch from one company to another, much thank to aggregating sites on the internet, where customers can compare prices between different companies. With the whole range of competitor's prices compared for the customer, brand becomes likely less important and price more important. It changes the behavior of customers and how and where they choose to make their business. With higher pressure on prices comes lower marginal and profitability. And here is where loyalty comes in. Keeping an existing customer requires usually less effort and money than acquire a new one. There is tons of literature on the subject, most claiming that it costs between four to ten times more to acquire a new customer than to retain one. Retention is more effective and has higher ROI, Return of Investment, than acquisition. It is therefore important to know your customers and how loyal they are to gain an advantage in the market. The information of customer loyalty applies to both which customers to retain and which to acquire, by pricing or targeting. Understanding customer behaviors can be extremely valuable and much can be gained by focusing on customers in segments or with certain properties, which tends to be loyal. So, loyalty is an important factor for the business and loyal customer is something a company wants in order to keep up with competitors and tough market conditions.

As an exception and clarification, loyal customer will not necessarily lead to higher profitability. For example, if the pricing is incorrect and unprofitable, a loyal customer will actually mean bad business for the company.

1.4 Retention programs

Länsförsäkringar is a federation that consists of 23 different regional companies, sharing the same brand. They share some of the functions, such as IT, national marketing and development teams. To meet the competitive market and to make customers more loyal, Länsförsäkringar developed a retention program. Retention programs are an effort on the behalf of a company to try to retain valuable customers and making them more loyal. Retention programs can be designed in many ways, but the most common one in the insurance industry is to discount certain products if the customer commits in one or several other products or services that the company has to offer. We say that the discount is given conditioned on the service or product. The conditioning terms for a discount differ between the companies, as well as the discount levels. The more committed the customer is the higher discount level he or she usually gets. From the company's perspective the aim is to increase the loyalty of customers, grow on certain products or services, diversify the risk and increase profitability. From the customer point of view it is also a good deal, since the customer has the option to get discounts and reduce the costs for insurance cover while enjoying the convenience of fewer financial providers. It is believed to be a win-win situation.

1.5 How?

Many factors influence customer decisions, so it is difficult to predict the actions regarding loyalty of a customer and the reasons behind. A way to learn what drives and makes a loyal customer is to build a model for the behavior. The model will try to find behavioral patterns regarding loyalty depending on the customer's properties, status, services, commitment and so on. We will especially study how a customer acts in a car insurance product, while being committed to other products and services, as in a retention program, and to see how different commitments influence their loyalty. The three commitment levels that will be studied are

- No commitment
- Partial commitment – Product and service A
- Full commitment – Product and service B

Full commitment means that a customer has more products and/or services than partial commitment. Whether the model measures the effect of lock-in, convenience, customer satisfaction or something else is for this objective irrelevant. The focus will be on measurable effects, not explaining reasons.

One can see loyalty as something that can be bought. If you give the customer something in return for being a good customer, the customer is also more likely to renew. The most common way this is done is by giving customers discounts, as in Länsförsäkringar's retention program. A higher discount leads to a more loyal customer; at least, that is what is to be expected. This paper especially aims to explore what is loyalty depending on commitment into other products and services and what is loyalty due to a discount. To do this we need to model the actions and reactions regarding loyalty of customers in car insurance.

1.6 Procedure

We will build the model from scratch, adding possible and eligible properties of customer and car, to see if they can explain the behavior of the customer with a car insurance contract. The properties that will try to explain the behavior is called explanatory variables.

In the first model, including only commitment as an explanatory variable, the commitment level will capture both price and non-price effects. This is to get a starting point for how much commitment will seem to matter and influence the choices of the customer. It is important to remember that this model will measure both the effect of commitment and price matters. We will after this learn and follow how the value of this variable changes as we build our loyalty model to include more explanatory variables, such as the discount level. This model will be called the Commitment Model.

Secondly, we will add all relevant and significant control variables into the model, on top of the first model with only the commitment variable. We do this in order to check if some of the effect that was captured in the Commitment Model actually was caused by something else. For example, suppose that all non-committed customers only were young persons, we would in the first model actually measure the effect of persons with commitment versus no commitment and young versus old persons. By adding the control variable age, we would measure the influence of the two explanatory variables separately. This model will be referred to as the Non-price Model.

To be able to separate the price and non-price arguments in the commitment level variable, there has to be a relevant and realistic way to model the effect price has on the outcome. All price issues might matter, so we will therefore discuss the price variables one by one and motivate the inclusion, except for the discount variable which will get a section of its own. This is referred to as the Price Model.

Finally we will build the complete model with discount and the eventual interaction between the explanatory variables. The full model will be called the Complete Model and will be compared to the previous model to see how much of the loyalty depends on discount and how much is commitment and to what extent they affect the measured loyalty. With the Complete Model the challenge will be to find a good way to measure the strength of the commitment such that we can compare it to the discount level in an intuitive way.

3 Loyalty model

3.1 Background

Loyalty models have been widely used for decades now, mainly within market strategies, where it is used to calculate the loyalty or duration of a customer. The purpose of building a loyalty model is to learn more about customers and their behavior; what drivers affect loyalty, who cancels their contracts frequently, at what rate and so on. Knowing different customer's behavior regarding loyalty can be used for several things; building retention programs, setting discount levels, scoring, targeting, pricing optimization, predicting, etc.

3.2 Model type

Measuring how long a customer is a customer as a time span, or duration, can be quite tricky. If you have the data from long time ago, it can now be out of date. If you are to use new data you have no way to know how people act after they have been customer a while. A common way to quantify and measure loyalty is to model the retention rate and then calculate the duration based on that rate. Retention rate is the ratio between the retained customers and the number at risk. Another way would be to call it renewal probability, since insurance contracts are renewed yearly and it is with a probability that the customer decides to renew or not. The renewal probability π will therefore be the measure of loyalty. A higher probability means a more loyal customer. And by modeling the renewal probability we avoid the problem of outdated or limited data.

Renewal probability should not be confused with renewal theory, which has a different approach, with time and amount as two variables. Since renewals always occurs once a year, there are no arguments to make a more complicated model than necessary, such as suggested by D. R. Cox (1962), even though including amount into the model is appealing, although out of scope here.

The statistical model we will use consists of a sequence of independent random variables Y_1, Y_2,…where Y_i has a binomial distribution with parameter π_i and n_i, such as

$$Y_i \sim Bin(n_i, \pi_i) \tag{1}$$

We will use the notation i for discussing both cells and customer level. In the case of Y_i meaning renewal or cancellation on customer level, as is the case with continuous explanatory variables in the model, n_i will be equal to one. When analyzing a single categorical explanatory variable, Y_i will refer to the grouped data in a specific cell i, where n_i is larger than one. It will be apparent what is intended in each section. For the understanding of the mathematical theory and assumptions regarding categorical variables it is helpful to sometimes group the data.

We want to model the probability π for a customer with a car insurance contract. A renewal model is simulating a binary outcome; either the contract is renewed or cancelled

$$y_i = \begin{cases} 1, & if\ i'th\ customer\ has\ renewed \\ 0, & if\ i'th\ customer\ has\ cancelled \end{cases} \tag{2}$$

Cancellation is the complement to renewal. The variable y_i is in this case a realization of the Bernoulli distributed random variable, since $n_i = 1$. It can take the values one and zero with the probability π_i and 1-π_i respectively. It can be written in the form

$$Pr\{Y_i = y_i\} = \pi_i^{y_i}(1-\pi_i)^{1-y_i} \tag{3}$$

With expected value and variance

$$E(Y_i) = \mu_i = \pi_i \tag{4}$$

$$Var(Y_i) = \sigma_i^2 = \pi_i(1-\pi_i) \tag{5}$$

Suppose we now divide the observations into groups, or cells, depending on driver age, car model, geographical zone, commitment and so on. The cell i has n_i observations. Let us assume that all individuals in a cell are homogenous, that is, has the same mean and variance. In short, customers with the same properties behave equally. We also assume independence between the cells. How a certain type of customer behaves does not affect another type of customer. If we are to use data from several different years we also need to assume time homogeneity, which is realistic since we only are using data from the last three years. Something that happened, or did not happen, last year, does not affect this year.

The probability function can be written as

$$Pr\{Y_i = y_i\} = \binom{n_i}{y_i}\pi_i^{y_i}(1-\pi_i)^{n_i-y_i} \tag{6}$$

y_i can be interpreted as the number of renewals from a total of n_i customers in cell i.

We want to estimate the renewal probability for every cell, π_i using regression analysis of the observations we have. Each probability π_i will be based on the explanatory variables x_i and the vector of regression coefficients β. The simplest model would be a linear function of the explanatory variables

$$\pi_i = x_i'\beta, \qquad \text{where } x_i' = (x_{1i}, x_{2i}, \dots, x_{pi}) \text{ and } \beta = (\beta_1, \beta_2, \dots, \beta_p) \tag{7}$$

A problem with this model is that there is no guarantee that the probabilities will stay between zero and one, which is of course a necessity if we are to work with probabilities. What are our options? Which model suits our purposes best? The two simplest, yet powerful and most commonly used models are the logit and the probit regression models. They differ in the assumptions of the errors. While the probit assumes normal distributed errors, the logit assumes logistically distributed errors. What does that mean? It means that the link functions of the two models differ. Using the above formula with the probit approach and the regression model will look like

$$\phi^{-1}(\pi_i) = x_i'\beta \tag{8}$$

ϕ^{-1} is the inverse of the Cumulative Distribution Function (CDF) of the standard normal distribution and called the link function in this case. If we are to use the logit model, the link function and the model will look like this:

$$log\left(\frac{\pi_i}{1-\pi_i}\right) = x_i'\beta \tag{9}$$

Which one is to prefer? The general opinion is that there hardly exists any difference between the results of the two. In the Hahn & Soyer paper “Probit and Logit Models: Differences in the Multivariate Realm” an interesting summary and new empiric evaluation is discussed. They find that model fit is improved by selecting logit for multivariate link function models. Even though it is arguing for a logit approach for our case, there is no rule when to use either one of them. If your variables are directly mapped to zero and one, as in a Bernoulli trial, then your outcome is binomial and logit would be preferable, since it also is the canonical link of the binomial distribution. Canonical link means that the link function is on a certain form which leads to some advantages and desirable properties regarding especially sufficiency and the derivation of the Maximum Likelihood Estimator. More on canonical links, see Agresti, 2003, p. 148-149. The logistic regression also has the advantage of being easily interpreted and almost always included in data and analyzing packages. If, on the other hand, we were to model something that we dichotomized into zero and one, for example high/low blood pressure and if we believed blood pressure to be normally distributed between people, then probit might be a better choice. As renewal is clearly a Bernoulli trial; either a customer renews or cancels, we will continue build the model using the logit approach.

But how did we get to the link function in equation (9)? Any mapping on the interval zero to one would actually do. We will construct the link function, η and go through the mapping, from zero and one to the entire real line, as we will encounter useful information that we will need in later sections.

We can move the boundary from zero and one by transforming the probability in two steps. We start with the upper boundary, by transforming the probability into odds

$$Odds_i = \frac{\pi_i}{1-\pi_i} \tag{10}$$

Odds are often used in the betting industry, but it also exists in mathematical and statistical tests and literature. They should not be confused. Odds, in the Swedish betting sense, are called European odds and are defined as a return from a betted amount. For example, a European Odds of 1.5 means that from a betted amount of 100kr, you get 150kr back if you win. You are plus 50 kr. It corresponds to a probability of 2/3 for a win.

The odds used in statistics can be interpreted as the probability of win divided by the probability to not win. For the probability example above, the European Odds of 1.5 corresponds to a statistical odds of 2, which in many ways make more sense. It is twice as likely to win as not to win.

With this transformation, odds can take all positive values. Secondly, we take the logarithm of the odds to calculate the log-odds, or logit

$$\eta_i = logit(\pi_i) = log\left(\frac{\pi_i}{1-\pi_i}\right) \tag{11}$$

This moves the lower boundary and η_i can take values ranging from minus to plus infinity, which is what we were trying to achieve. As the probability π_i is approaching zero, η_i reaches minus infinity and when π_i is approaching one, η_i reaches plus infinity. The inverse transformation is

$$\pi_i = \frac{e^{\eta_i}}{1+e^{\eta_i}} \tag{12}$$

So, probabilities between zero and one maps one to one to the entire real line of numbers.

Instead of using the simplest model in equation (7), we take the *logit* to be a linear function of the explanatory variables.

$$\eta_i = logit(\pi_i) = x_i'\beta \tag{13}$$

We have now a generalized linear model with a binomial response and a link logit with explanatory variables x_i and the vector of regression coefficients β.

See Rodriguez, G. (2007), p. 6-9 on how to solve the numerical equations for the maximum likelihood estimation of the β's as it is not within the scope of this paper. We will instead use available software packages for this. Since different tools have different strengths, the model will be implemented in base SAS and Towers Watson's Emblem. Using two tools has the advantage of categorical graphical analysis and the whole range of statistical tests with continuous variables.

3.3 Perfect renewal

The aim of building the renewal model is to be able to predict and learn how customers behave in a renewal situation. To do this we observe how the probability of renewal varies with certain explanatory variables. The most common way to model renewals is to use a logistic regression, as we discussed in section 3.2. For the logistic regression we needed to assume time independence, cell independence and cell homogeneity. Another assumption we need to make is that all customers are expected to act and behave similarly regarding price and commitment no matter where they live, which regional company they are member of, how old they are, which car they drive and so on. In short, people are equally price sensitive and act equally when committed into more products or services concerning their insurance contracts. Commitment is not in every regional company connected to a certain discount level. Some companies give no discount at all for partial or full commitment, while others have the whole range of discount levels. The assumption covers also that customers act similarly when exposed to a discount and that the discount levels are known to the customer.

If the assumption would not hold, we would either need to test commitment and price for interaction with all other variables or build 23 separate models. Not only would that be time consuming, but really difficult to find reliable estimates for the variables, due to lack of data. It would be hard to draw any conclusions at all.

Thankfully, the assumption is intuitively realistic as long as we can exclude observations where we might expect a different behavior. To help this assumption pass we will construct a "perfect" renewal.

In this particular case, where the main focus is to study price and commitment, we do not have to care so much about the predictability of the entire portfolio of customers. We can therefore severely simplify the model to only be valid for certain customers, certain contracts and certain events. If we were to predict a complete portfolio and the loyalty of customers, we would have to take all possible events into account in the model.

Let us start with the definition of contracts and renewals. An insurance contract is an agreement between a customer and an insurance company, regulating insurance cover for an object, in this case a car. Loosely speaking, a renewal is a customer having an insurance contract of a car that is continued after the contract's end date, initiating a new yearly contract between the insurer and the insured. The renewal happens passively without any action from the customer as long as premiums are paid. If a customer wishes to discontinue the contract, the action has to be made by the customer.

A perfect renewal occurs when the contract is renewed the day after the end date of the previous period with no changes to terms or conditions in the product. This means that all observation with changes to the contract or renewing at another date will be discarded. The latter usually happens when the contract has changed in any way. Example of this type of observations that will be excluded is whenever someone changes cover, address change initiating a new contract, insuring a new car and so on.

A perfect cancellation is the complement to a perfect renewal. A perfect cancellation occurs when the cancellation date is exactly the same date as the end of the contract period. Therefore we can exclude all cancellations due to sold or crashed car, address change and so on. A cancellation at the end of the period can be triggered by premium raises, bad experience or general dissatisfaction, or similarly.

The number of cars a customer has insured is a strong explanatory variable for a high renewal probability. To model multiple cars require a different definition of what a renewal is. For example, how to deal with a customer that at the start of the period has two cars insured and at the end only one? A similar case is for the cars not currently in traffic. To avoid these types of situations and to keep modeling a perfect renewal, we will exclude customers with multiple car insurance contracts and cars not currently in traffic. Vehicles need also be under normal external market conditions, which make some models inappropriate to analyze, such as sports or veteran cars. Completely new vehicles are subject to guarantees from the manufacturer for the casco part. These guarantees are usually active in two to three years. In order to avoid the renewal transfer from guaranteed casco to casco that the customer pays themselves, these cars are excluded. On the other end there is a certain bump that occurs at the vehicle age of 25. To avoid modeling this bump, vehicles older than 24 will be excluded from the subset. More special cases that will be discarded are contracts with odd discounts, unusually high premiums or price adjustments. They are few and for the sake of homogeneousness, they will also be discarded from the analysis.

Even though income, tradition and cultural behavior might differ slightly between geographical regions in Sweden, it is considered to be immaterial for the basis of a loyalty model. People are the same, but market strength in different regions might not be the same. Market strength is although more connected to history and strategy of a specific company rather than the people living there, and that effect will be captured in the company variable that we will study more in later sections.

Young people are in a completely different economic situation, so they will therefore be excluded from the study. At the age of 26 most people owning a car are also on payroll, which make them eligible for study. Young people also tend to be more mobile, which might separate their behavior from the rest of the population regarding commitment. Another completely different situation occurs when people become pensioners, or senior citizens. They are no longer on payroll, have ample of time to optimize their financial situation. We can also see that there is a bump at the year of 65, where people in Sweden normally retire from work. That will be the upper limit. So, for the assumption to hold, young and old people need to be excluded from the analysis.

Another thing that has proven to alter the elasticity is whenever a customer has had a claim recently. A customer with a claim tends to accept an increase in the premium more often than a customer who had had no claim recently. To keep the price sensitivity homogenous, observations with a claim last year will be removed. Customer that is changing their commitment status is also removed, just to make sure we do not deal with price changes that are expected beforehand.

After filtering all above, we are left with approximately one million observations from a total of 3 million. The observations come from the years 2011 to 2014. There is a small variation between the years. It is not assumed to have anything to do with the change in behavior of customers or change in market conditions, but rather an unexplained variation.

Subset

Argument	*Filter*
Owner age	26-64
Vehicle age	4-24
Nr. of contracts per customer	1
In trafic	yes
Temporary discounts	No
Claim last year	No
Made changes to contract;mileage, cover, etc	No
Switched car	No
Made changes in commitment factor	No
Price change due to individual adjustment	No

Figure 1: Filters to create the perfect renewal

Data will be structured as one observation per row in the way the table in figure 2 shows, although with a lot more columns than just commitment. It has a lot of advantages by doing so. We can compare goodness of fit between different models, we do not need to regroup data each time we test new variables and continuous variables do not need to be rounded up or truncated, and so on.

Observation i	Renewed y_i	Commitment x_{1i}	Commitment x_{2i}
1	1	0	0
2	0	1	0
3	1	0	1
4	0	0	0
5	1	1	0
6	0	0	1
7	1	1	0
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
n	1	1	0

Figure 2: Structure of the data set

3.4 Expectations

What we expect to find by first building a partial model and then build, brick by brick, price elements into the loyalty model is that the coefficient for the commitment factor will change for each element, to finally drop when discount level enters the model. If the commitment factor remains stable even in the full model, the conclusion will be that price does not matter for a customer with commitment. On the other end of the scale, if the commitment factor reduces to zero, or at least if a hypothesis test for the coefficient is not significant anymore, we will conclude that price is everything and commitment has nothing to do with loyalty and there will be no point in having retention programs. Since we expect to have collinearity between the two variables of interest, we will have to explore how to deal with such issues as well.

3.5 Odds Ratio

In a multiplicative model such as this the odds ratios plays an important role when understanding a logistic regression model. For every variable, one of the levels is chosen to be the reference, usually the most stable level where we have much data. We can abandon that standard and choose the most logical as a base, for example for the commitment variable (as we have seen before the non-commitment is a more natural way to relate to the other levels of commitment).

A simple model with only one explanatory variable will look like this

$$\eta = log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \quad (14)$$

β_0 is called the intercept and equals η for the reference cell, in this case the no commitment level. A categorical variable with k levels is represented in the computation with $k-1$ binary explanatory variables being either zero or one, as this is an ANOVA type model. In this case $k = 3$. η for the partial commitment is $\beta_0 + \beta_1$ and full commitment is $\beta_0 + \beta_2$. β_1 represents the change in the logit scale when going from no commitment to partial commitment. This can be quite difficult to interpret on the logit scale. If we on the other hand exponentiate it we get the odds.

$$\frac{\pi_i}{1-\pi_i} = e^{\beta_0+\beta_1} \tag{15}$$

Odds are the probability of an event divided by the probability of the event not happening. The odds for renewal in the reference cell (for no commitment) are e^{β_0} and the odds for the other two levels are $e^{\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}}$, where either x_{1i} or x_{2i} is zero. The effect in odds of having no commitment changes with the ratio of e^{β_1} for partial commitment and e^{β_2} for full commitment. The ratio is not the total odds but how the odds are changing with a chosen factor or variable. For example, the odds ratio OR_2 for the partial commitment equals

$$OR_2 = \frac{Odds_2}{Odds_1} = \frac{\frac{\pi_2}{1-\pi_2}}{\frac{\pi_1}{1-\pi_1}} = \frac{e^{\beta_0+\beta_1}}{e^{\beta_0}} = \frac{e^{\beta_0+\beta_1}}{e^{\beta_0}} = e^{\beta_1} \tag{16}$$

And how is this to be interpreted and why is it better than comparing β's?

If there is no difference in the outcome of different commitments, then the odds ratios will be one for every level since $\beta_1 = \beta_2 = 0$. An odds ratio of two for the partial commitment would mean that the odds are two times more for partial commitment than for no commitment. Also the full commitment relates to the reference cell, such that an odds ratio of two means the same thing as for the partial commitment and there is no difference in renewal between partial commitment and full commitment. If no commitment had a probability of renewal of 2/3 (odds of 2.0), then the partial and full commitment will have a probability of 0.8, from the definition of odds

$$Odds_i = \frac{\pi_i}{1-\pi_i} \tag{17}$$

If you are used to working with loglinks and multiplicative models for Poisson or Gamma distributed variables, working with odds ratios is a natural thing to do, since the relativities are used in the same way as the odds ratios. Even if you are not used to working with neither loglinks or odds, one can more easily transform the odds to probabilities than β_{ik}.

3.6 Duration

If the renewal probability is assumed constant, then the distribution of the customer's life span, or customer duration, *X,* is geometrically distributed with mean

$$Duration = E(X) = \sum_{i=0}^{\infty} i\pi^{i-1}(1-\pi) = \frac{1}{1-\pi}\ , 0 < \pi < 1 \tag{18}$$

This is only true if the π is constant and $i \to \infty$, which is of course never the case for an insurance contract. As time goes by, the probability of renewal will change. Even if π might stay relatively constant, people cannot live forever, so the duration will only serve as an approximation. The smaller π, the better approximation, since problem arises with long duration, cause it is there life expectancy and other changes to the probability really impacts. Before that a perfect renewal is quite close to the general renewal for a newly acquired car. Claims are quite rare, changing cars happens every five to ten years, we have only included customers in the study with enough life expectancy and so on.

The duration is only used for a pedagogical purpose, since it is easier to see the effect of different renewal probabilities related to a time-span rather than a probability.

3.7 Building the model

When building a regression model we want it to be parsimonious. Parsimonious means that the model accomplishes the desired level of explanation with as few variables as possible. We also want the model to predict well. There is no complete method to achieve these goals. One can not only rely on a single measure. Every decision that includes another explanatory variable should be made with care and caution. Sometimes none of the measures are applicable as they test for different things. Variables might score low on goodness of fit tests but are at the same time important for future predictions. Because even though dependencies are weak, in terms of statistical measures, it does not mean that it is not important. When only a few explanatory variables are of special interest it is critical to study and learn as much as possible about their correlation and eventual interaction. An example of this is how price sensitive customers are. We might have scarce data to learn from, yet we need to know, or at least see a tendency of how customers react to price changes in order to predict well. In these cases it can be valuable to be able to look at the data using some kind of graphical tool. We will use all the measures and tools at our disposal to build the best possible model for renewal.

Since the objective of this paper is to separate a certain effect from a variable by adding another, we know we can expect a certain degree of collinearity. We will test for the severity of the collinearity and discuss the possible problems that will arise due to this.

3.7.1 Predictive power

The first important thing one need to consider regarding predictive power is how reliable the variables are. The model might fit well to a set of observations, but when it comes to predicting the future, reliability is very important. When including an explanatory variable in a model, you will see to which degree it affects the outcome by looking at the β or preferably the odds ratio. But you also need to know how reliable this estimate is, i.e. the null hypothesis for the variable. The null hypothesis can be formulated as: the explanatory variable that is studied has no effect on the renewal probability. It can be done by looking at the confidence intervals or making a hypothesis test for the estimator. It will be conducted at a significance level of $\alpha = 0.05$. A Wald test is a standard hypothesis test, very easy to compute and always included in software packages, that compares each explanatory variable and will be made and considered for all variables as the model grow to include more variables. Since we might be dealing with collinearity and interactions it is important to test all of the variables every time we add a new factor and not only when it is added. For a coefficient to qualify as an explanatory variable we want a rejection of the Wald test at a significance level of $\alpha = 0.05$ at all times.

The accuracy of models can be assessed in several ways. The two most common are by calibration and by discrimination. Calibration relates to how well predicted probabilities match to observed, while discrimination concerns how well a model can separate those who have higher and lower probability of the event of interest. The measure that is most used for calibration is the coefficient of determination, R^2. For the logistic regression, we cannot interpret R^2 in the way we do ordinary least squares, since there is no minimizing the variance, but an iterative way to reach the maximum likelihood. They are therefore called pseudo-R^2, because they remind us of R^2 and range from zero to one. There are several variants of pseudo-R^2, such as McFadden's, Efron's, Cox-Snell's and so on. For a more complete view on different R^2-measures for logistic regression models, see Hosmer, Hosmer, Le Cessie & Lemeshow (1997). As it is really difficult to interpret, understand and find satisfactory levels for the different R^2-measures it will not be used for model choice.

Another approach is the Hosmer-Lemeshow test. Expected values for each observation are calculated in the data set and ordered from highest to lowest and grouped into ten groups in approximately equal size. For each group the observed and expected numbers of counts are registered. Pearson chi-square is then applied to compare expected and observed. This solution has been criticized for being random in its outcome depending on what number of groups one chooses and will therefore not be a test to depend upon, see Allison in the Statistical Horizon blog (2013).

Since none of the calibration measures reaches a satisfactory level of reliability or explainability we will rely upon discrimination measures instead. For discrimination, the area under the Receiver Operating Curve, also called concordance statistic or C-statistic, is most commonly used. It measures how well the model discriminates and differentiates the predictions. It is often used when the ordering of the predictions is more important than the probabilistic accuracy. The C-statistic is a strong test for the predictive power for binary outcomes regarding differentiation and discrimination. It reminds us of the Kolmogorov-Smirnov test (KS test), as they are both ordered. KS tests the maximum vertical distance between two distributions and answers the question (in our case): are the data from a logistic distribution? Since we are confident in the strict binary outcome of the independent variable, we will not use KS test.

The C-statistic is closely related to the Receiver Operating Characteristic curve (ROC). It is actually the area under the curve. Then, what is the the ROC-curve? The easiest way to describe it is with a graphical approach.

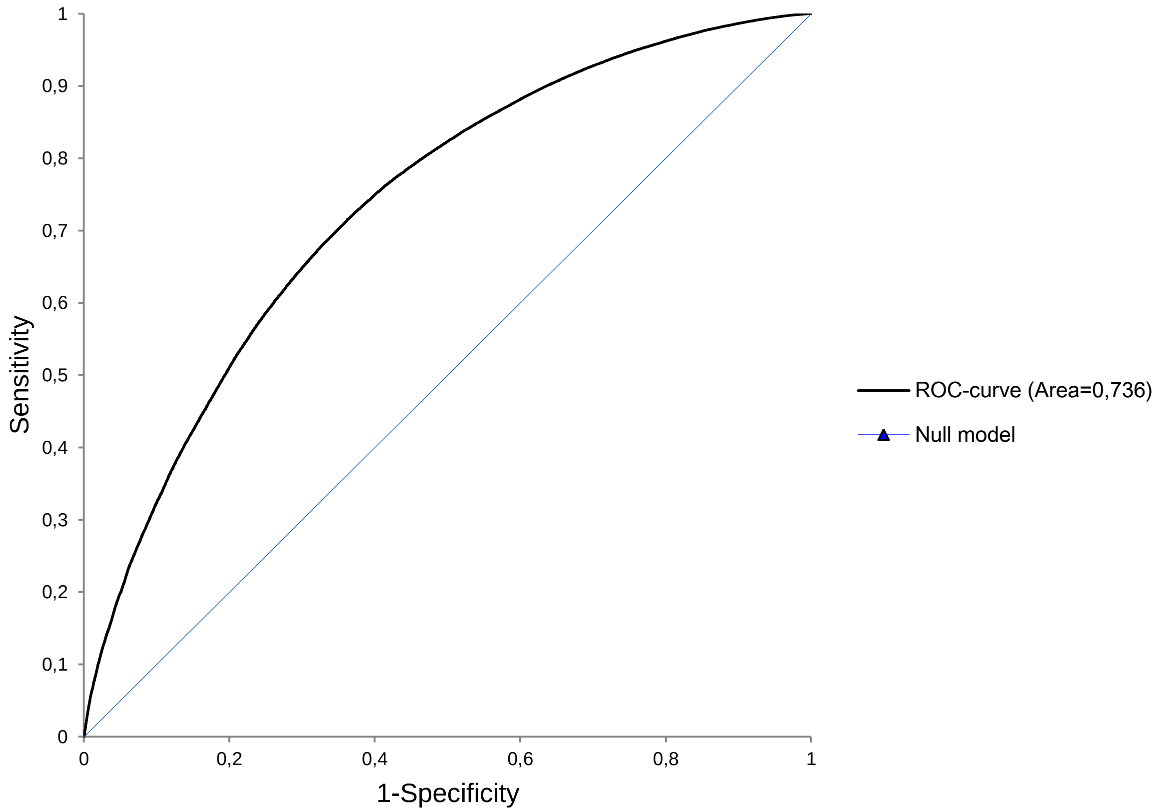

Figure 3: Area under the curve and the C-statistic

What we need to do to compute the ROC-curve, is to estimate the renewal probability for every observation we have in our data set with the current model. After this we sort the observations in ascending order of the estimated renewal probability and divide it into small levels or segments. For every level of estimated renewal probability, we count the true positives and the false positives. The accumulated true positives, also called the sensitivity, measures the accumulated proportion of the correctly identified. For example, if 50% of the total cancellations occurred for 10% of the highest estimated cancellation rates, then 0.5 is the sensitivity. The accumulated false positives are the accumulated proportion of the incorrectly identified. For example, if 5% of the total renewals occurred in the 10% of the highest cancellation rates, then the false positives is 5%. This is often referred to as 1-specificity, since the specificity is the complement, namely the true negatives. Going through all levels and accumulate, we get the x and y-axis of the ROC-curve. We can now construct the graph in figure 3. Interpreting the graph above, the larger area under the curve the better the model discriminates. The C-statistic is the area under this curve, sometimes also called AUC. It is a measure between 0.5 and 1 (without ever reaching 1). Models are typically considered reasonable when the C-statistic is higher than 0.7 and strong when C exceeds 0.8 (Hosmer & Lemeshow, (2000), p. 162).

Suppose all of the observation had the same estimated probability, then the curve would be spread along the diagonal, i.e. the zeros would be equally spread over the whole x-axis. In the graph it is called the Null model. On the other hand if the model could predict exactly in the correct order, then we would approach y=1 as fast as we can count the zeros. For example, if one tenth of the observation has cancelled, then y=1 for all x>0, see the red line in the graph below. The C-statistic will play an important role in building a renewal probability model.

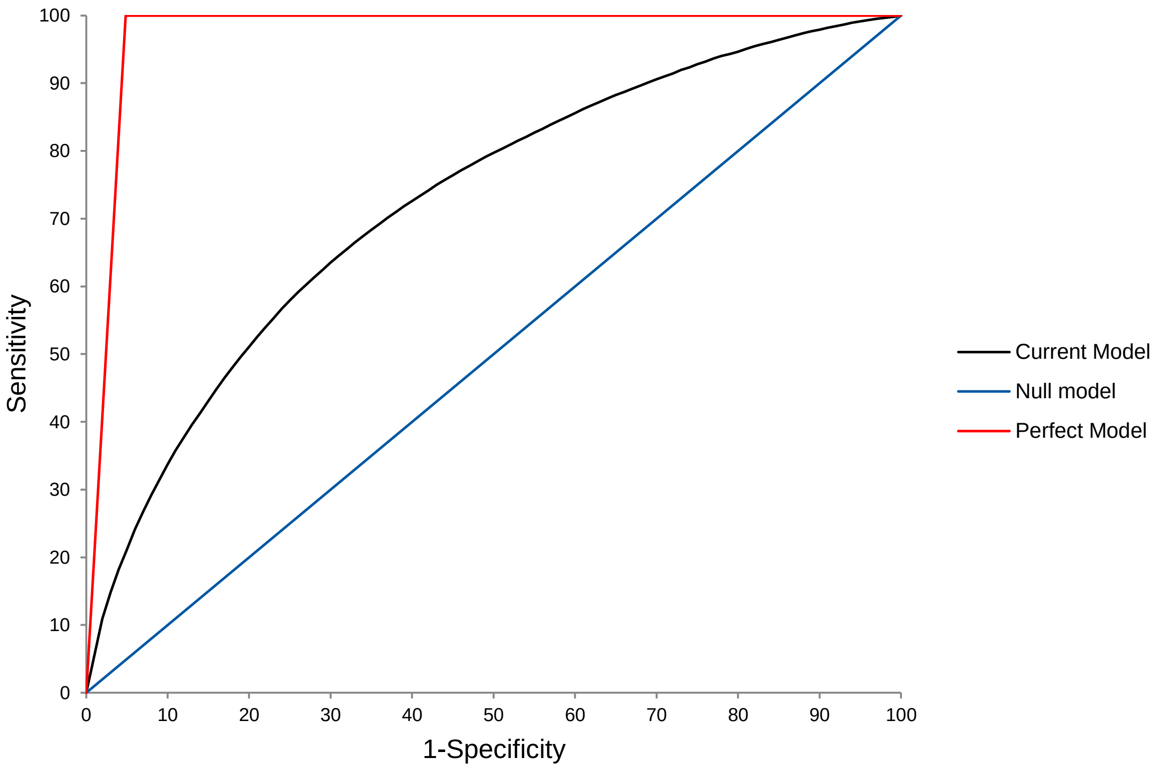

Figure 4: Area under the curve and the perfect model

3.7.2 Parsimony

Parsimony means that the model needs to be well balanced between complexity and goodness of fit. Adding another variable will always result in an equal or better fit, but the question is if the change in model fit is statistically significant? To answer that question we will perform a hypothesis test of the goodness of fit between two models. Since we are building the model by adding one variable at a time, it is reasonable to perform some kind of verification with every added variable that the model has improved. And by improved, we mean in a parsimonious sense.

There are a few goodness of fit-tests that can be done to compare models, but all have disadvantages and no perfect solutions exist. The most popular assessment techniques for comparing models regarding goodness of fit between models are the Wald test, likelihood ratio test and the Akaike validation.

The Wald test is an approximation of the likelihood ratio test and lacks the precision when comparing complex models with many variables. But for testing every variable standalone it serves its purpose.

The likelihood ratio test is a standard test and commonly used for comparing models. A model fit almost always improves when the model gets more complex (more variables) and the test is answering the question: Is the model fit significantly better? The test is based on the deviance statistic. The deviance is defined as

$$D(\mathbf{y}, \widehat{\boldsymbol{\mu}}) = 2[\ell(\boldsymbol{y}) - \ell(\widehat{\boldsymbol{\mu}})] = 2\sum_{k=0}^{n} \left\{ y_i log\left(\frac{y_i}{\widehat{\mu}_i}\right) + (n_i - y_i) log\left(\frac{n_i - y_i}{n_i - \widehat{\mu}_i}\right) \right\} \qquad (19)$$

y_i is the observed and μ_i is the fitted value for the *i*'th observation. If we calculate the deviance statistics for two models, we can make a likelihood ratio test to compare if the fit is significantly better or not. The difference in the deviance for the two models follows a χ^2-distribution with *k* degrees of freedom, where *k* is the difference in degree of freedom for the two models. It is a powerful test and applicable even for continuous variables and ungrouped data, as the deviances are subtracted. The likelihood ratio statistic is

$$\Delta G^2 = D(\mathbf{y}, \widehat{\boldsymbol{\mu}})^s - D(\mathbf{y}, \widehat{\boldsymbol{\mu}})^r \qquad (20)$$

s represent the smaller model with less parameters than *r*. This is under the null hypothesis that the smaller model is correct.

Akaike has of course some attractive properties with the number of parameters in the model acting as a penalizing term, balancing the goodness of fit between simple model and overfitting. With continuous variables and binary data, as in our case, the penalization becomes rather meaningless since the deviance becomes so large compared to the number of parameters in the model, which makes the penalization too small to really affect the deviance. The pure deviance and the likelihood ratio test will therefore be used to compare models.

3.7.3 Collinearity

Even though commitment is a central definition and the discount level a local decision, there is a correlation between the two variables. This is because we know that many of the companies condition their discounts on some of the commitment, such that it is likely that whenever a person is committed, they also have a discount. Whenever this is the case, it might lead to unstable estimates of the variables. In fact, we expect this to happen, although, we do not know to what extent.

In the GLM computation, the order of the explanatory variables does not matter, and therefore there will be no parameter being computed “first” or “last”. The variables will be computed at the same time and both variables will be affected by the collinearity. If there are collinearity issues, we are likely to see an increase in the Wald’s p-value of one of the added variables, meaning larger confidence limits.

A solution to most collinearity issues is to get more data. There is of course exceptions to this; for example if a variable is a linear function of another variable. We know that this is not the case, since some of the companies do not have any discounts conditioned on commitment. The collinearity problem arises as the Wald’s p-values rise high, the variables vary more and uncertainties increase. To be sure we are within boundaries between the commitment level and discount level, we will check the Variance Inflation Factor (VIF).

$$VIF = \frac{1}{1-r^2} \qquad (21)$$

where r^2 is called the coefficient of determination. For this case, we can actually use the standard definition of r^2 as this coefficient comes from a linear regression on that variable alone using all the other variables. For example, a VIF of 1.5 means that the variance is 50% larger than it would be without the correlation.

At what quantities does the VIF pose a problem? Most expert on the subject call it a collinerarity problem when VIF>10. Allison (2012) believes it becomes a problem as VIF reaches 2.5.

4 Non-price Model

The partial model is supposed to include all arguments that are relevant and significant apart from price arguments that can describe the renewal probability π_i. We will start building the model with one of the two main variables, the commitment variable. We do this to get a feel for the data and how well this variable predict and fit to the data. After that we will include control variables. Control variables are variables included in the model that are known to affect the dependent variable, but are not of importance to the main goal. They are needed so that we do not capture irrelevant effects that originate from factors such as age, car model or something else that may affect the renewal probability.

4.1 Commitment model

For a model with only the commitment argument as a variable for the renewal probability, the model will look like

$$\eta_i = logit(\pi_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \qquad (22)$$

The coefficients β_1 and β_2 explains the effects of the commitments. It is modeled as two variables taking the values zero and one. π_1 is the probability of renewal for a car insurance contract where the subject has no commitment in Länsförsäkringar apart from the car insurance, π_2 the probability for a contract where the customer has partial commitment, and π_3 when the customer has full commitment.

We will choose no commitment as the reference cell. That way it is easier to see the effect of each commitment level has on the outcome. β_0 is the intercept in this model, and the probability for a person in the reference cell is therefore

$$\pi_1 = \frac{e^{\eta_1}}{1+e^{\eta_1}} = \frac{e^{\beta_0}}{1+e^{\beta_0}} \qquad (23)$$

Before we look at coefficients, odds ratio and renewal probabilities for the model with only the commitment variable, we will study some of the output statistics from the regression model, as we have described in section 3.7. We compare the commitment model M_c with the null model. The null model is a model with no explanatory variables and only an intercept, β_0 from equation 22.

Statistics	M_{null}	M_c
C-statistic AUC	0.5	0.629
Deviance $D(y, \hat{\mu})$	419 547	406 185
LRT ΔG^2		13 362
Degrees of freedom		2
Prob < χ^2		<0.0001

Figure 5: Statistics to compare models

The C-statistic, also called the Area Under the Curve, is a measure of how well the model can discriminate the contracts using the predicted renewal probabilities. It is ranging from 0.5

(no discrimination) to 1.0 (maximum discrimination). The deviance, $D(y, \hat{\mu})$ is a quality-of-fit statistic for models, for which the Likelihood Ratio Test (LRT) is based upon. The LRT is comparing two models and testing whether the improvement of the more complex model is significant or not. The difference between the deviances of the two models follows a chi-squared distribution with k degrees of freedom and the null hypothesis will be rejected if p-values are less than 0.01. These statistics will be shown for every new model we build. All of the measures and tests in figure 5 are thoroughly explained under section 3.7.

We can clearly see that the commitment variable is an argument that has a lot of predictive power with a C-statistic of 0.629, which is far better than any random model, including the null model, which has 0.5. And this was achieved by adding only one parameter. The likelihood ratio test (LRT) is rejected, which means that the improvement of fit is significant in the new model compare to the null model, according to the likelihood ratio test, so we can be confident that commitment plays a big part in a customer's choice to renew or not. These statistics will be displayed for every model. Wald's null hypothesis on the other hand, will be tested for every explanatory variable and model, but will not be displayed in a table. A rejected null hypothesis will be a requirement for all variables in any model.

If deviance continues to drop enough such that the likelihood ratio test is passed and the C-statistic keep increasing enough we will go on including factors. We will of course not include and test nonsense factors. There have to be relevance to all variables included. Also, we should be prude in adding more factors to a model. "Things should not be multiplied unnecessarily", is the approximate meaning of Occam's razor, quoted from Ohlsson & Johansson (2010), p. 62. It should be interpreted as, if possible, keep the model simple and parsimonious.

Next we will look at the variable itself and not the model.

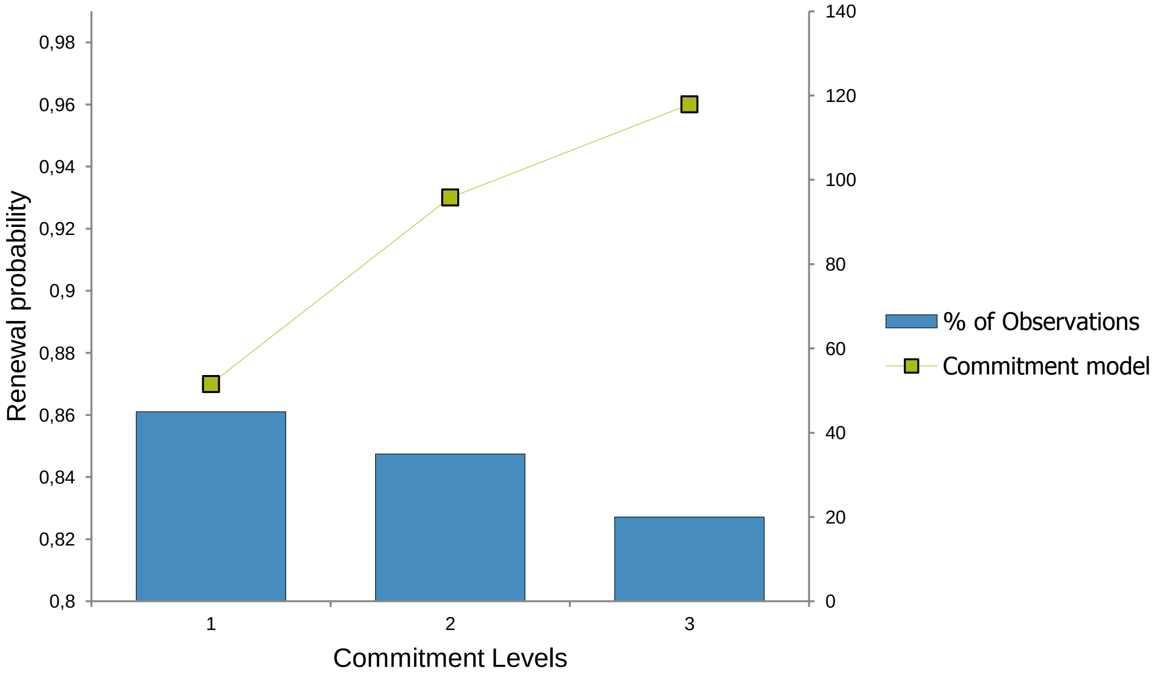

Figure 6: Renewal probability for commitment levels

The first interesting thing to notice is that when we delete all "imperfect" observations, we are left with a quite high probability of renewal. It is of no importance, since we are trying to figure out the difference in behavior between customers and to do that we needed to get rid of disturbing exception. Most of the cancellations occur for young people and when a current customer switch car, as we have discussed under section 3.3.

Explanatory variable	Cell i	Coefficient β_i M_{null}	M_c	Odds ratio OR_i M_0	M_c	Probability π_i M_0	M_c
Intercept		2.10	3.26			0.89	0.87
Commitment	1		0,00		1.00		0.87
	2		0.62		1.86		0.93
	3		1.41		4.10		0.96

Figure 7: Estimates of the null and commitment model

We will be focusing on the odds ratios and not the probability or the coefficient from now on as we explained in 3.5, since it is much easier to interpret. We see that the odds ratio more than doubles as we go from partial commitment to full commitment. It is a substantial change going from no commitment to a commitment, but before we go further we need to see the confidence levels for the commitment variable.

Odds Ratio Estimates

Variable	M_c	Point Estimate	95% Wald Confidence Limits	
Commitment	2 vs 1	1.859	1.822	1.896
Commitment	3 vs 1	4.096	3.885	4.319

Figure 8: Confidence limits of the commitment levels

The confidence intervals are really small and that indicates a good fit. We will not disclose the Wald test but only the confidence limits, if it is not of special importance, but we will make sure that we are at a significance level of 0.05 at all times, if not stated otherwise. Remember that, at this point, the commitment variable will capture some, or all, of the effect from the lower price that customers with commitment often have. And maybe it also captures some other effect that we do not know of yet.

4.2 Non-price control variables

For the non-price model we want more variables in order to capture effects that come from other than price factors. These variables are called control variables. They are not of any interest apart from removing their effect from the equation. A typical example would be age. Even though we have excluded the really young and really old, there is still a difference between ages as renewal concerns. We do not want the commitment or discount level to pick up that effect. The model will then look like equation 24 which is a development from equation 9.

$$\eta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_m x_{mi} \qquad (24)$$

Where $\beta_m x_{mi}$ are the variables for each cell *i*. This is the non-price model, where no price components are included. The first index indicates the order of variables and the second index the cell number.

Engagement Years is the time, in years, that the customer has had a car insured in Länsförsäkringar. Drivers Age, Vehicle Age and Engagement Years are modeled as continuous variables. All these variables will be referred to as control variables and have one degree of freedom each. We will not display or discuss the control variables individual influence or their linearity as our focus is on commitment and discount.

Below is the non-price model, M_{np}, written in a more readily way.

$$M_{np} = Base + Commitment + Drivers\ Age + Vehicle\ Age + Engagemant\ Years \quad (25)$$

Below are the output statistics comparing M_{np} to M_c.

Statistics	$\boldsymbol{M_c}$	$\boldsymbol{M_{np}}$
C-statistic AUC	0.629	0.671
Deviance $D(y,\hat{\mu})$	406 185	400 797
LRT ΔG^2		5 388
Degrees of freedom ΔDF		3
Prob < χ^2		<0.0001

Figure 9: Statistics comparing commitment and non-price model

We can see in the figure 9 that we have increased the predictability somewhat and have at the same time passed the likelihood ratio test that is testing whether M_{np} has significantly better fit than M_c.

Odds Ratio Estimates

Variable	$\boldsymbol{M_{np}}$	Point Estimate	95% Wald Confidence Limits	
Commitment	2 vs 1	1.905	1.832	2.001
Commitment	3 vs 1	4.031	3.804	4.401

Figure 10: Confidence limits of the commitment levels

The odds ratios for the commitment levels stands firm even with the introduction of the control variables. It is a good sign indicating stable coefficient. We can no longer compare probabilities as the probabilities depend upon several other factors that now are included. That is why, from now on, we will only talk of odds ratios.

One can be tempted to conclude that commitment explains all the loyalty and start building retention programs based on this information. But that can be a dire mistake. There is a discount that comes with the commitment in most cases. Meaning, we are actually buying their loyalty. The question remains, how much loyalty are we buying and to what cost? To explain this, we need to include all relevant and explaining factors regarding price matters. Even though we are only interested in the discount level, we need to include the control variables for price. Also, there are more issues to account for when it comes to price. That is why inclusion of price components into a regression model is worth a section of its own.

5 Complete model

There are many different aspects of how a price may affect the renewal probability. In fact, there is no limit of how complicated pricing in insurance can be and the real difficulty lies in how well one can simplify to make it manageable. In this paper we will divide price matters into four categories: Local strategy, market competition, price change and discounts.

Since we merely are interested in how the pure discount affects the renewal probability, we see the other price variables as control variables. We want to remove any possible disturbance from them so that we can observe the pure effect from the discount. The control variables in this section require a bit more explanation than in the previous section, since price is so much more complicated and versatile.

5.1 Price control variables

5.1.1 Company variable

Länsförsäkringar consists of 23 different companies placed in different regions of Sweden, which acts independently of each other regarding most strategies and price issues. They partially share brand, yet there are still large differences between the levels of loyalty. This variable is supposed to capture the regional strength of brand and price strategy as well as their market competition. Therefore a variable for company is essential to being able to model the loyalty realistically for each region. The variable will be called Company variable and is modeled as a categorical variable.

5.1.2 Premium level

When people choose insurance company for their car, price is a really strong factor for which insurance giver the customer chooses. When it comes to renewal of contracts, the relationship to price is not as strong and factors like commitment becomes more important. Since we also have reduced the data set to include as homogeneous observations as possible, a fair assumption is that Länsförsäkringar is equally competitive in all of the chosen segments, such as customer age, brands, car model and so on. Competitors are likely to have the same information, or at least equivalent, regarding claims cost, and pricing is done in the same manner with a multiplicative tariff as a base. The only segments that truly can be observed with different renewal probabilities are between price levels. A higher premium leads to a lower renewal probability. This can be due to market competition, but is more likely to have to do with the incentive a customer paying high premiums have to find cheaper alternatives compared to a customer paying less. Either way, this variable will capture both effects and will serve its purpose as a control variable. The premium level will be modeled as a continuous linear variable.

5.1.3 Price change

When a customer is exposed to a price change, it triggers a reaction from the customer. It can increase or decrease the probability of renewal, depending on the price change. This is usually referred to as price elasticity or price sensitivity. As above, customers are more sensitive to price when acquiring a contract than renewing and it will become too hard to model patterns on individual level, so we will assume that all customers are equally price sensitive. For more on price elasticity, see Varian, Hal R. (2006), p. 266-287.

To get the best estimate of the price change, one would prefer to randomly test the sensitivity. Since all 23 companies of Länsförsäkringar regulate their own premiums, as well as their discount levels, the premium changes can be assumed independent between companies and individuals. This makes the elasticity easier to observe and fairly good, especially compared to a case with one company making all the price adjustments for everyone every year. The price change will be modeled as a change in monetary amount, rather than a change in a percentage. Since premiums can vary from 2 000kr up to 25 000kr, a percentage change will prove to be more complex to model. And also people tend to relate and react more equally to for example 1000kr than 10 percent, regardless of income or social status. The price change will be modeled as a linear continuous variable. It is not a fact that a linear variable would fit well to observations of how customers react to price changes of an insurance contract. It has to be studied and tested before one can assess which way is the most appropriate. As can be seen in figure 11, a polynomial of degree one fits well enough and there is no need to investigate it further.

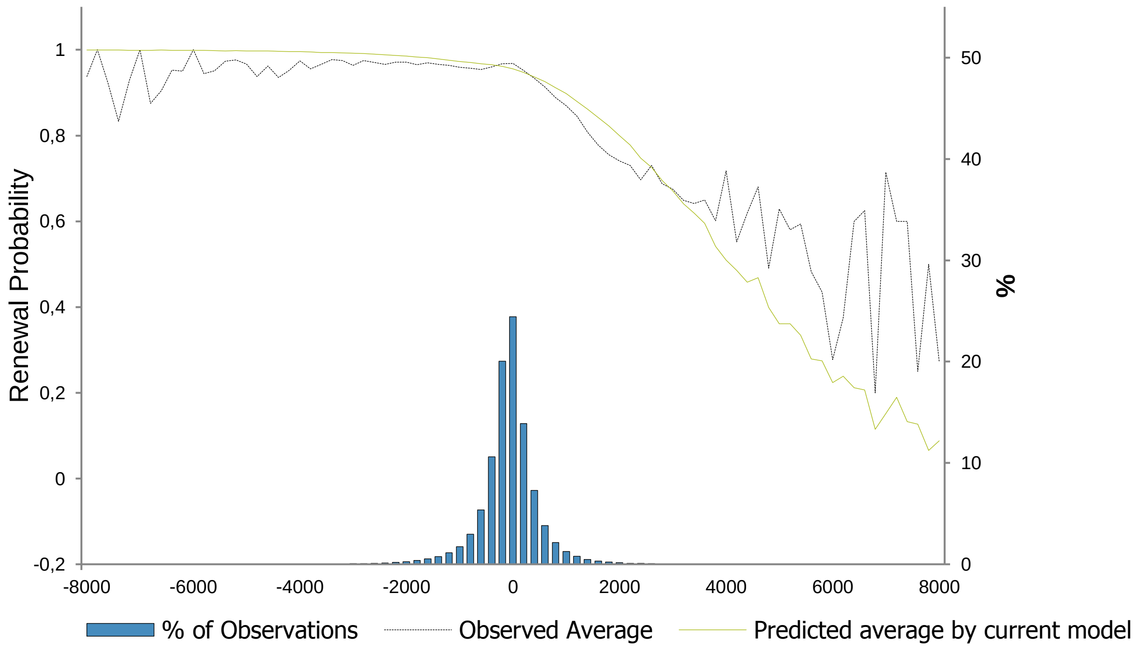

Figure 11: Price change between contract periods

The model with the control variables for price looks like:

$$M_p = Base + Commitment + Drivers\ Age + Vehicle\ Age + Engagemant\ Years + Company + Price\ Level + Price\ Change \quad (26)$$

Adding the control variables for price gives the following output statistics, all of which has significant Wald's test.

Statistics	M_{np}	M_p
C-statistic AUC	0.671	0.729
Deviance $D(y, \hat{\mu})$	400 797	384 104
LRT ΔG^2		16 693
Degrees of freedom ΔDF		23
Prob < χ^2		<0.0001

Figure 12: Statistics comparing non-price and price model

We have now reached a satisfactory C-statistic greater than 0.7 and the improvement of fit is significantly better as the likelihood ratio test is passed. We have a model that has potential of separating well and predicting good. Every model with an added variable has successfully passed the LRT and not only the comparison between M_p and M_{pc} as in the table above.

Odds Ratio Estimates

Variable	M_p	Point Estimate	95% Wald Confidence Limits	
Commitment	2 vs 1	1.787	1.722	1.857
Commitment	3 vs 1	3.744	3.443	4.062

Figure 13: Confidence limits of the commitment levels

It seems that the control variables for price have captured some of the effect that the commitment factor formerly had since the point estimate has dropped significantly from 1.905 and 4.031 respectively for the partial and the full commitment. It happened as the variable of price change entered the model, so there was probably premium decreases occurring for committed customers in some companies, making the commitment higher before the price change was added. It is a good example of how important control variables can be when assessing effects from variables. If we were to use this model to predict future renewals, adding price change to the model would be even more important.

5.2 Complete model

5.2.1 Adding discount level

Discounts will be modeled separately from the price. A natural way is to see the discount as a ratio, or percentage, paid premium divided by full premium. That is also how it is displayed and marketed to the customer, the so-called communicated discount. Often the customer reacts more positively to a discount percentage than a competitive price, since the customer usually does not have a full view of what is a competitive price and what is not. The discount level will be modeled as a first degree polynomial which the graph below helps to motivate. The fitted average is with the variable included in the model. The linear trend is clear, but since different companies within Länsförsäkringar have different discount levels, the slope will become jumpy. The company variable tries to explain this but falters where the data is scarce, for example where the discount level is 13% and 17%. These discount percentages are very rare among the companies in the Länsförsäkringar group.

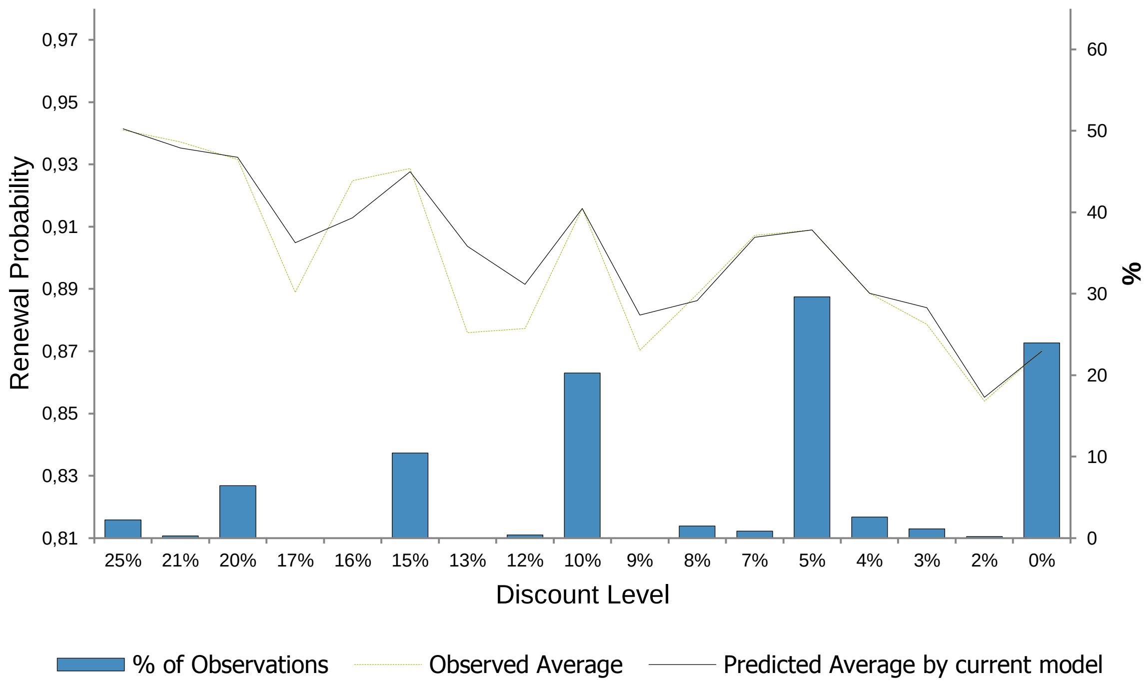

Figure 14: Discount and renewal probability

The output statistics from the model with the discount level, M_{pd} is shown in table below.

Statistics	M_p	M_{pd}
C-statistic AUC	0.729	0.735
Deviance $D(y, \hat{\mu})$	384 104	381 032
LRT ΔG^2		3 072
Degrees of freedom ΔDF		1
Prob < χ^2		<0.0001

Figure 15: Statistics comparing price model and price discount model

We notice that we are reaching convergence for the C-statistic. As we have seen in earlier sections this is not surprising. If we run a regression with only the discount level as variable we would have a C-statistic of 0.631, which is a strong C-statistic for only one variable. When the model already has many explanatory variables, the C-statistic is not so useful anymore. The improved fit is significant and we can note that discounts have a positive effect on the renewal.

Odds Ratio Estimates

Variable	M_{pd}	Point Estimate	95% Wald Confidence Limits	
Discount	Continuous	529.8	466.1	608.9

Figure 16: Confidence limits of the discount variable

The discount variable is continuous which means that the coefficient β of the discount variable should be multiplied by the actual discount, between zero and one, and then taken the exponential in order to see the effect as an odds ratio. It becomes quite hard to interpret the meaning of such odds ratios, which is why we only show it for the confidence limits. As can be seen, it is a strong explanatory variable with lots of predictive power.

Odds Ratio Estimates

Variable	M_{pd}	Point Estimate	95% Wald Confidence Limits	
Commitment	2 vs 1	1.583	1.543	1.624
Commitment	3 vs 1	2.210	2.055	2.374

Figure 17: Confidence limits of the commitment levels

Now we have a real change in the odds ratio of the commitment factor. The odds ratio has dropped from 1.787 to 1.583 for partial commitment and from 3.744 to 2.210 for full commitment. Confidence limits indicate only a small increase in variance. We were expecting collinearity between the two variables and that really proved to be true. The odds ratio for a fully committed customer almost halved when discounts entered the model!

In cases of severe collinearity, the variances of the coefficients become really large. We cannot see any signs of that when looking at the confidence limits above. To be absolutely certain that the model will not suffer from collinearity issues we calculate the Variance Inflation Factor. Correlation between commitment and discount is approximately 0.5 and the VIF is below 1.38 which indicates that this should not pose a problem in the model, see section 3.8.3.

5.2.2 Adding interaction

We now turn to see if we have specified the model concerning commitment and discount level correctly in terms of polynomial order and interactions. We test the second and third polynomial for the case of a quadratic or cubic linearity for the discount Level. It contributes nothing to the Deviance nor for the C-statistic so it will not be implemented. Higher degrees will contribute even less, so the investigation ends here regarding higher order polynomials. We also test for interaction between Discount Level and Commitment. For the model with the interaction term we have a significant improvement of fit which motivates to explore the interaction further.

Statistics	M_{pd}	M_{pc}
C-statistic AUC	0.735	0.735
Deviance $D(y,\hat{\mu})$	381 032	380 907
LRT ΔG^2		125
Degrees of freedom ΔDF		2
Prob < χ^2		<0.0001

Figure 18: Statistics comparing price discount and complete model

Even though as far as deviance and C-statistic concerns, the contribution is small. The likelihood ratio test is significant and encourages us to include it. P-value is also really low and within confidence limits. In cases like these it is valuable to be able to look at the data.

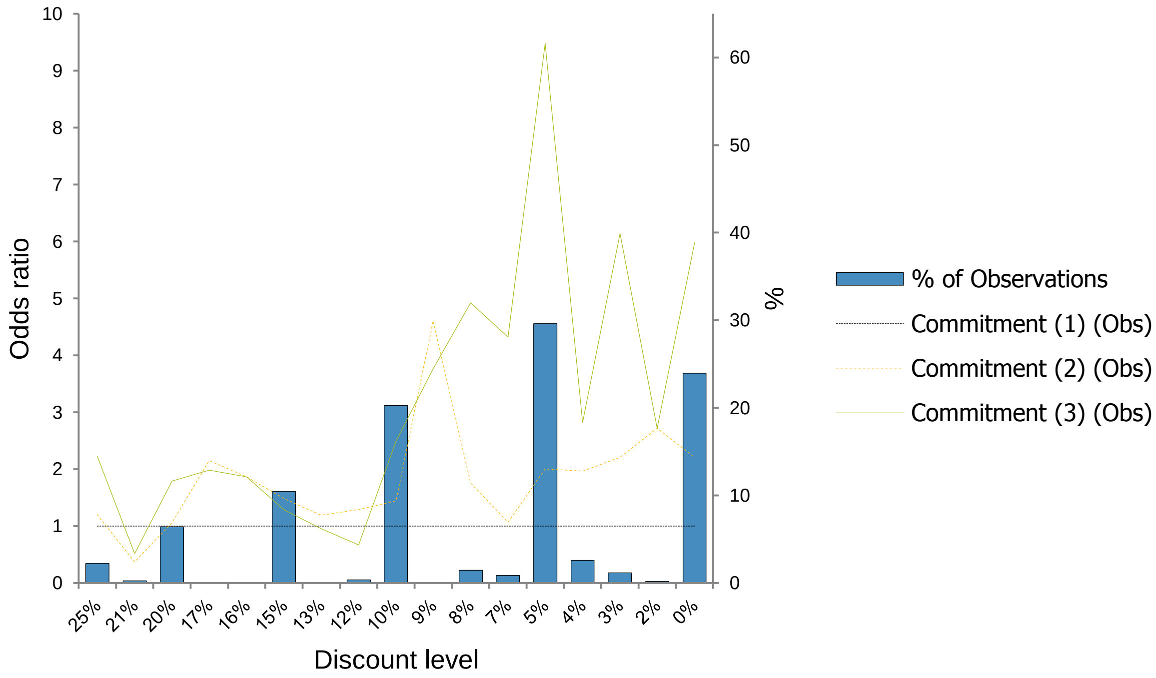

Figure 19: Observed odds ratio and discount level related to no commitment level

Figure 19 of the odds ratios and the discount levels requires some reflection. We can clearly see that customers with high commitment and low discounts are more likely to renew than their counterpart, a customer with no commitment and low discount. But we can also see that commitment matters less for high discounts than for low. The commitment curves seem to converge as the discounts get higher, see figure 19 to the left. It makes sense. As a discount becomes really high, satisfaction or dissatisfaction matters less, as long as the customer is happy with the low price. The cost of cancelling the contract is too big and once a customer has moved all of its products to a chosen insurance giver, a slightly smaller or larger discount don't matter as much as when you are not as committed.

For the understanding of customer behavior this interaction means a lot, which strengthens the means to include it. Since the aim of the paper is to separate loyalty from commitment and discount, this is an important information and should be included as an explanatory variable.

We now have a full model for the "perfect" renewal, M_{pc}.

$$M_{pc} = Base + Commitment + Drivers\ Age + Vehicle\ Age + Engagemant\ Years + Company + Price\ Level + Price\ Change + Discount + Commitment * Discount \quad (27)$$

The typical and average discount for the partial commitment is somewhere around 5%, corresponding an odds ratio of 2.0 from figure 20. It is approximately the same odds ratio as in the commitment model M_c and works as an acknowledgement that we have understood the relationship between the two variables correctly.

Explanatory Variable	Cell *i*	Odds ratio *Commitment=1*	*Commitment=2*	*Commitment=3*
Discount Level	0%	1,00	1,59	2,61
	5%	1,20	2,02	3,08
	10%	1,68	2,58	3,64
	15%	2,37	3,28	4,30
	20%	3,33	4,18	5,08
	25%	4,69	5,33	6,00

Figure 20: Odds ratios of commitment and discount

5.2.3 Stressing the model

We would like to test the stability of the model by calculating the ROC-curve (the C-statistic) once again. In section 3.7.1 we displayed the ROC-curve for the dataset that we also used for building the model (sometimes called the training dataset). We will now test the model with another dataset. We will use the same coefficients from the model M_{pc}, but on the next year's data, 2015. We do this to estimate the renewal probability of each observation in the new dataset. We order the observations by ascending predicted renewal probability and calculate how many cancellations of the total we captured in every percentile and plot it in figure 21.

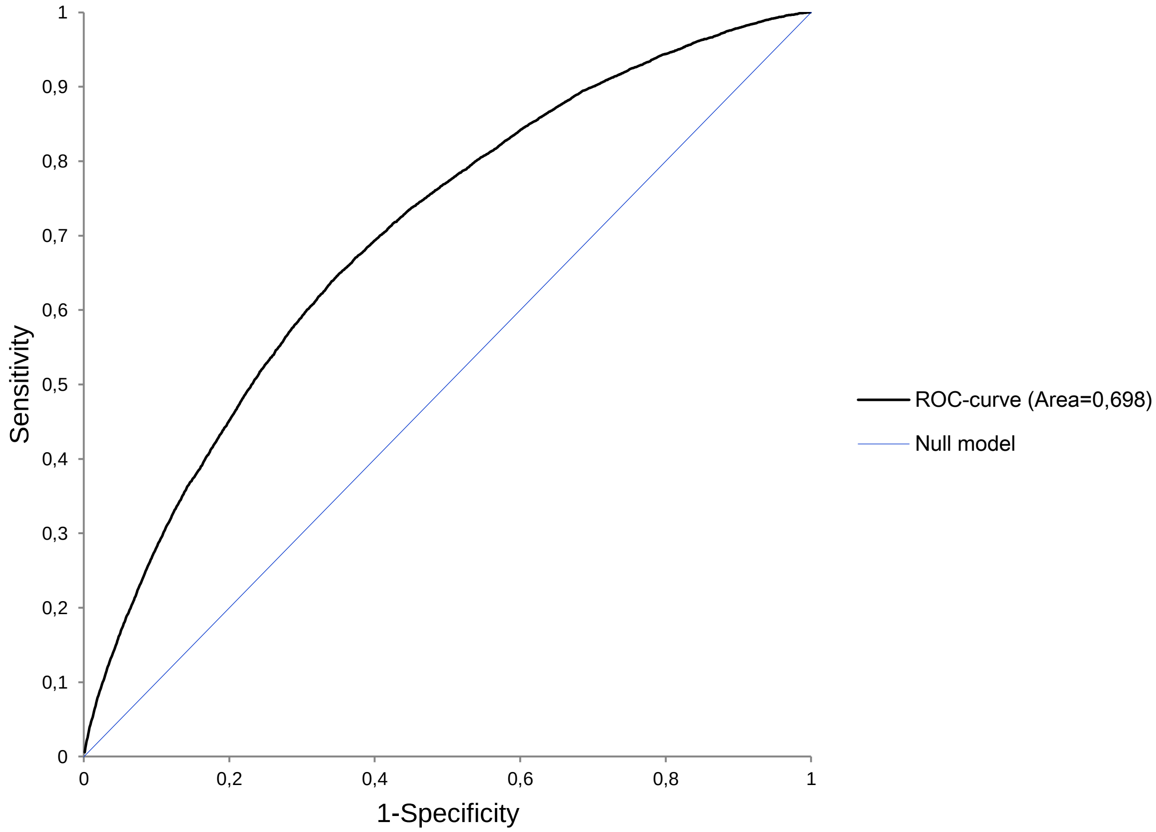

Figure 21: C-statistic when stressing the model with a new data set.

Compared to figure 3 we can hardly distinguish any difference between the ROC-curves. This is a good sign! It means that this model has some stability and is able predict quite well the order of the renewals and cancellations. The C-statistic has dropped from 0.735 to 0.698, but the number is still a strong discriminative accuracy. A small drop was of course to be expected, since you can never beat a dataset of which the model was built.

This test will not catch a change in the level of estimates, meaning that the predicted renewal probability can be far from actual, yet this test will show good numbers as long as the predicted renewal probability are in the correct order. Depending on what you are doing with the model, this usually is not a problem. If the aim is to make exact predictions of the future, another measure to complement the C-statistic might come in handy. If we are to assess relationship between the variables or select the most loyal customer, this test is fully satisfactory.

If we would have tried to build the model, M_{pc} using data from a single company, the uncertainties would become larger, as can be seen in figure 22. We call this model M_{pc^*}. The VIF increases from 1.38 to 1.60 which is still not troublesome. But what is problematic is that there would not be enough significance evidence of including the discount level and the confidence intervals would become really wide. We would probably have coefficients changing between years because of the uncertainties in the estimates of the coefficients. It is therefore recommended that in order to assess the relationship between variables, more data than from a single company is desirable.

Odds Ratio Estimates

Variable	M_{pc^*}	Point Estimate	95% Wald Confidence Limits	
Commitment	2 vs 1	1.780	1.559	2.032
Commitment	3 vs 1	3.762	2.351	6.019

Figure 22: Confidence limits of commitment for a single company

6 Conclusion

6.1 Approach

We wanted to investigate how much of the loyalty in a retention program that is price related and what depends on other factors. To do this we studied the retention program in Länsförsäkringar. A typical retention program gives discounts conditioned on commitment into other products or services and so does Länsförsäkringar. For example, if a customer acquires a certain service in Länsförsäkringar, the premium for the car insurance contract will be discounted by a specified percentage. But how much of the increased loyalty is due to a competitive price and how much is convenience, satisfaction, lock-in effects and so on?

Länsförsäkringar consists of 23 sovereign companies sharing the same brand. Every company sets their own discount level and on what products and services the discounts are conditioned on. There are similarities of how they choose what to condition on, but also differencies. That makes it easier to see how different discount levels affect the loyalty. The commitments that will be studied are centrally determined and not necessarily connected to a discount, although many are. The three commitment levels that will be studied are

1. No commitment
2. Partial commitment – Product and service A
3. Full commitment – Product and service B

To be able to achieve the goal of separating discount and commitment we defined loyalty as renewal probability and built a binary regression model. See section 3 for definition and decisions regarding model type. The model is supposed to capture how different properties of customers and contracts, called variables, may affect the choice to renew. To build a regression model we used a mathematical technique called GLM.

To investigate the relationship between the two parameters of interest, we needed to build a full model for the loyalty, starting from scratch. For every added variable we used three measures to validate the model; the Wald's test for the null hypothesis of every explanatory variable, the C-statistic for the predictability and the likelihood ratio test for comparing current model with the previous and smaller model. In section 3.7 there is more theory and the motives behind the selected measures.

We started with a model M_c, which included only the commitment variable and then extended it to a model M_{np} with more so called control variables. Thereafter we added price control variables to the model M_p and finally we included the discount levels and an interaction between the two variables of interest M_{pc}. The model reached a C-statistic 0.735 which is above the desired limit as discussed in section 3.7.1.

To see how a variable affects the loyalty, we looked at the odds ratio of the variables and its supplemented confidence limits. In a stable model with a lot of data and no collinearity, parameter estimates are expected to vary less. As we expected collinearity, we needed to check for large changes in the odds ratios. We are specifically interested in how the odds ratio of the commitment variable changes as we add variables to the model, to learn as much as possible about the commitment variable before discounts enter the model. Finally we needed to verify if there was any interaction between the two variables.

6.2 Results

In the table below we can see how the odds ratio of commitment changed as we built our model to include all eligible and tested explanatory variables.

Explanatory variable	**Cell** i	**Odds ratio** M_c	M_{np}	M_p	M_{pc}
Commitment	1	1.000	1.000	1.000	1.000
	2	1.859	1.905	1.787	1.583
	3	4.096	4.031	3.744	2.210

Figure 23: Odds ratios of the different models

The dramatic change in odds ratio for the commitment variable occurs when the discount level enters the model, as was expected. The effect that we saw in the models without discount as an explanatory variable was picked up by the commitment variable. It means that the variables are collinear.

In most cases collinearity does not pose any problem as long as confidence limits are satisfactory. For this case it was, but to be sure we computed the Variance Inflation Factor, *VIF*. As VIF for the variables were within comfortable boundaries, see section 3.7.3. we concluded not to worry further and keep an eye on the confidence limits.

After testing successfully that the model with the interaction term is better than the one without, we composed figure 24 to see the relationship better.

Explanatory Variable	**Cell** *i*	**Odds ratio** *Commitment=1*	*Commitment=2*	*Commitment=3*
Discount Level	0%	1,00	1,59	2,61
	5%	1,20	2,02	3,08
	10%	1,68	2,58	3,64
	15%	2,37	3,28	4,30
	20%	3,33	4,18	5,08
	25%	4,69	5,33	6,00

Figure 24: Odds ratios of commitments and discounts

What the interaction term does is tilting the discount curve to become steeper for the no commitment level and flatter for the full commitment level. In the next section we will describe what this means.

6.3 Interpretation

The commitment and the discount variable have a lot of explanatory power. The small confidence limits strengthen this result. What it means is that they both have an effect on a customer's loyalty. A car insurance customer with discount is more loyal than one without. Whether customers have compared premiums with other competitive companies or they just trust that a premium with a discount is competitive is irrelevant. Discounts matter. We have also shown that having the car and product and service A in Länsförsäkringar makes a customer more loyal than one without. If the customer also uses Länsförsäkringar's product and service B, the customer is even more loyal. Possible explanations to the behavior can be that a larger part of the fully committed customers are satisfied customers that have good experiences with the company, compared to the no committed customers, and that they therefore wishes to stay longer. Another can be the convenience a customer experience when having many financial services in one place. Yet another is called lock-in effects, meaning that if the customer wants to change, for example some services to a competitor, the customer loses discounts on other products making it much more expensive to do so. Regardless, it is a fact that being committed to product and service A or B makes a customer more loyal.

We could see a large drop in odds ratio for the fully committed customer as the discounts entered the model. It turned out that the high loyalty related to full commitment, was really due to large discounts. It is not that the fully committed customers are especially discount driven or sensitive to prices, it is merely that large discounts have a large effect, with or without commitment. And large discounts are more common for fully committed customers.

As said in the section 6.2, a certain convenience effect can be seen for higher commitment levels. It can be interpreted as for every percentage spent on discounts, on the odds ratio scale, the effect is higher on the no committed customers than on the fully committed.

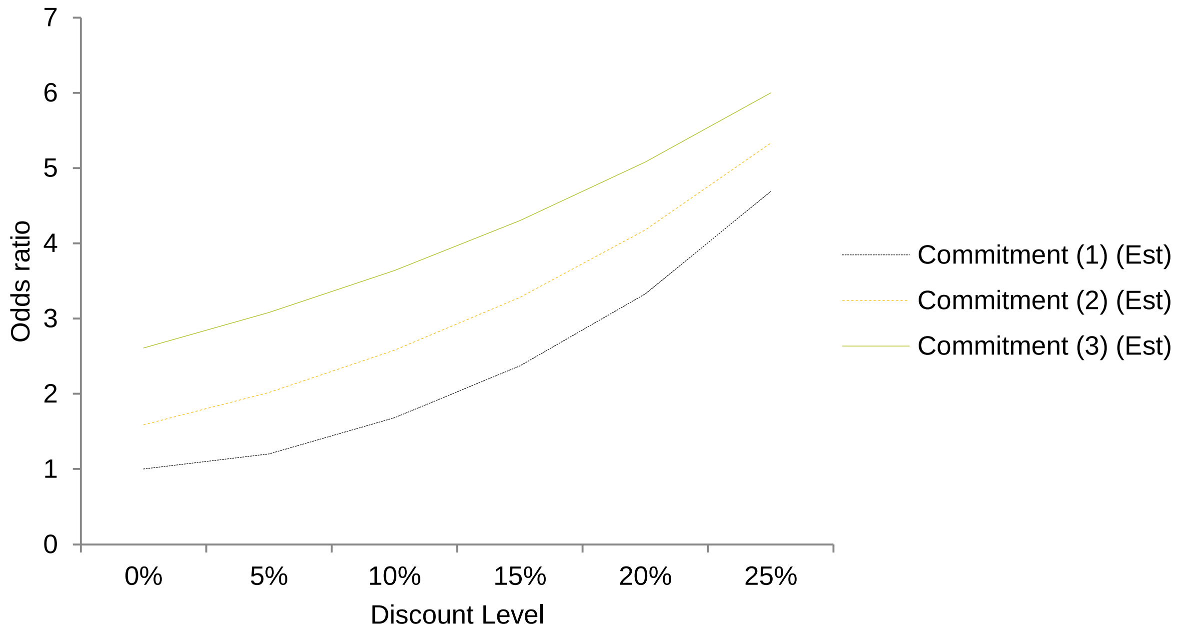

Figure 25: Odds ratios for commitment and discounts

We can see in figure 25 that discounts become less important the higher commitment. The curve for the fully committed customer at high discounts is not as steep as for the other commitment Levels. This means that if we want to work with discounts effectively, on the odds ratio scale, it is better used on customer with no commitment than on fully committed customers, especially for high discount levels. Then of course there is the problem of attracting new customer to the program, which shall be discussed in 6.6.

If we look at the odds ratio for commitment and discount level graphically, as in figure 25, we can take another approach: A customer that has a product and service A with Länsförsäkringar and 1% discount on the car is as likely to renew the car insurance as a customer with no commitment and a discount of 10 %, everything else equal. To make a customer with no commitment stay as long as a fully committed customer with a discount of 15%, we need to discount the contract 23%.

6.4 Simplification

Sometimes it can be hard to understand odds ratios, so we always have the possibility to convert it to a probability or even an approximated duration. Since the model consists of several explanatory variables, most customers will have different predicted renewal probabilities in the model because of their age, how long they have been a customer and so on. We have to decide at what probability level we would like to look at. Suppose we choose a customer with no commitment and no discount and suppose that he or she will have a renewal probability of 0.5. Every second year, this customer will cancel the insurance contract. Let us make an example using a six faced die. In our case it will mean that a roll of 1-3 will result in a cancellation and 4-6 a renewal of the contract.

To help understand the odds ratios for the commitment variable an easy comparison can be made. If a roll of 4-6 would result in a renewal for a car insurance customer without commitment and discounts, then 3-6 would be a renewal for the same customer with partial commitment in Länsförsäkring and 5% off the car insurance premium in discounts (corresponds to an odds ratio of 2.0). For the fully committed customer with 19% in discount, 2-6 would mean a renewal (corresponds to an odds ratio of 5.0).

Some prefer relating loyalty to duration. We can convert the probabilities to expected durations, which can be a bit more intuitive to handle, especially if we are trying to comprehend what a difference in renewal probabilities will mean in real life. The renewal probability of 0.5 in the example above corresponds to a duration of 2 years. More details on how to compute durations in 3.6. The fourth dot on the die, meaning the partial commitment with 5% in discount, will increase the duration to 3 years. Full commitment with a discount of 19% increases the duration to 6 years.

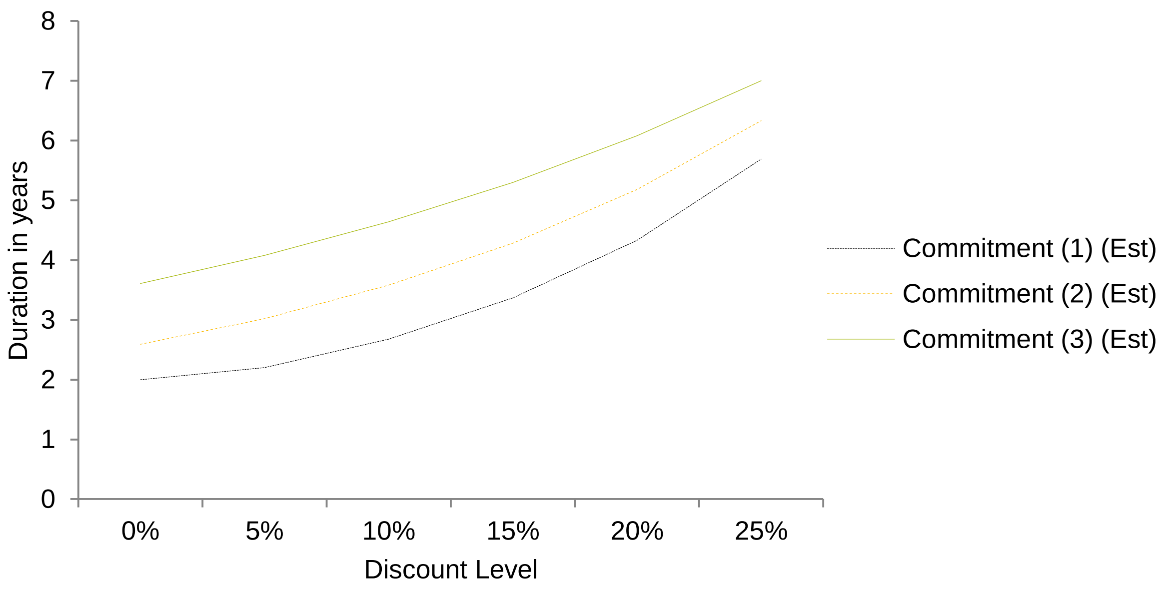

Figure 26: Duration for Commitment and Discounts

From the example above, we can roughly conclude that for a car insurance customer in Länsförsäkringar with full commitment and 19% in discount, we get almost two years extra for the commitment and another two years for the discount, on top of the basic two years. For the partial committed we get half a year from the commitment and another half from the discount.

6.5 Exceptions

We saw in the first model that we built M_c, that the three levels of commitment in renewal probability for a perfect renewal were 0.87, 0.93 and 0.96. It corresponds to the odds ratios of 1.0, 1.9 and 4.1, where discounts and other control variables are hidden within those numbers. It is severely larger than the probabilities in the six-faced die example above. The difference between the three level does not seem so high. If we instead approximate it to duration for the three commitment levels it would become 8, 14 and 25 years. On this scale the differences are huge. These durations are of course theoretical approximations, since there is no consideration taken to life expectancy, changes in life situations, claims occurring, changing of cars, etc, as we discussed in section 3.6. That is why the simplification with the six faced die is quite a realistic example. It shows us that we can increase the expected number of years a customer will stay using retention programs, it is just a matter of how much the company is willing to pay in discounts or converting customers to commitments.

Since we do not have data of discounts larger than 25%, it would be unwise to extend the curve, even if we could. There might be something that alters the relationship as we move towards higher discounts than 25%.

It is important to know that this is the case of pure renewals. In many cases there are special circumstances that affect a customer's decision. See section 3.3 for more details on the assumptions regarding perfect renewal.

6.6 Applications

We have up to this point not mentioned growth, as it is not the scope of this paper. In theory, the most profitable thing to do would be to make a customer commit without giving discounts. The customer would have an increased duration at no cost. Maybe it should be possible to take away discounts from the currently committed customer without losing them all to competitors, but it will probably be impossible to get new customer and the portfolio of committed customers will decline.

A retention program that conditions discounts on other products or services serves two purposes; to increase the loyalty on the discounted products and to grow on the products. Let us start with first one.

We can see a retention program as a way to enhance the loyalty effect. Like a boost for the discount. Instead of just randomly give away discounts, hoping customer to stay longer, the insurer makes a deal with the customer; if the customer invests, or commits to more products or services, they get a discount. It is an agreement between the customer and the insurance giver. This will make the effect larger than without the agreement. The knowledge of how and which effect that can be expected from customers can be used to set desired discount levels to reach certain loyalty levels or vice versa.

While an increased commitment only has positive effects on the profitability for the given product, giving away discounts has both positive and negative effects. Here one would want to know the cost for making discounts and for converting a customer to become committed to compare them with one another to find the most effective way to gain loyalty. Creating discounts is administratively relatively cheap compare to converting a customer. On the other hand, once a customer is converted to a commitment it is cheaper compare to what a discount costs. But making customers more loyal is no problem if there is no constraint on how much money we can spend. If we start by looking at the actual discounted product isolated, we need a way to measure loyalty of a customer in a monetary value to be able to compare it to the cost for the discounts. A typical approach is to measure profitability after financial principles, as a life-time value (LTV) or embedded value (EV), in order to get the effect from renewal probability into the profitability measure. Both are based on financial theory and how to measure future cash flows as a net present value. For each year, from now until the profits are estimated to be either too insecure or immaterial, the expected profit from a certain customer or contract is calculated and then discounted with a specified discount rate. Both approaches are basically calculating the same thing, but have separate application scopes. While LTV are founded in marketing for scoring of customers, EV was developed for the balance sheet in life insurance. More on mathematical approaches and its application in Berger & Nasr (1998) and Dwyer (1989). For details on the continuously developing Embedded Value, see CFO Forum (2009).

We can with a net present value approach go further and recommend a discount level for each of the different commitment levels based on the profitability instead of the loyalty, which is much more attractive way. We can also find the most optimized way to keep a customer by comparing giving discounts and converting customers to become committed with the cost to do it.

Up to now we have only seen what we can do for the actual customers and contracts and how they renew. If we could build another model for acquiring and attracting new customer, depending on commitments and discounts we can see the full picture. We could then base all decisions on loyalty and profitability.

7 References

Allison, Paul D. (2012): *When Can You Safely Ignore Multicollinearity?*, Statistical Horizon Blog (www.statisticalhorizons.com/multicollinearity)

Allison, Paul D. (2013): *Why I Don't Trust the Hosmer-Lemeshow Test for Logistic Regression*, Statistical Horizon Blog (www.statisticalhorizons.com/hosmer-lemeshow)

Agresti, Alan (2003): *Categorical Data Analysis.* John Wiley & Sons.

Berger Paul D. & Nasr Nada I. (1998): *Marketing models and applications,* Winter, Journal of Interactive Marketing, Volume 12, Issue 1, p. 17–30

Cox, D.R. (1962): *Renewal Theory,* Methuen Ltd,

CFO Forum (2009): *MCEV Principles & Guidance, MCEV Basis for Conclusions,* (www.cfoforum.nl)

Dwyer, Robert F. (1989): *Customer lifetime valuation to support marketing decision making, Journal of Interactive Marketing,* Volume 11, Issue 4, 1997, p. 6–13

Hahn, Eugene D. & Soyer, Refik (2005): Probit and Logit Models: *Differences in the Multivariate Realm, home.gwu.edu*, Salisbury University & The George Washington University, USA.

Hosmer, D. W. & Hosmer, T. &Le Cessie, S. &Lemeshow, S. (1997): *A comparison of goodness-of-fit tests*, Statistics in medicine, vol. 16.

Hosmer, David W. & Lemeshow, Stanley (2000): *Applied Logistic Regression,* 3rd Edition. New York: Wiley.

Ohlsson, Esbjörn & Johansson, Björn (2010): *Non-life Insurance Pricing with Generalized Linear Models,* Springer.

Rodriguez, G. (2007): Lecture notes, chapter 3: *Logit Models for Binary Data*, Princeton University, p. 6-9.

Varian, Hal R. (2006): *Intermediate Micro Economics,* 7th ed. W. W. Norton & Company.