

Advanced lifetime modelling to value customers in a multiproduct environment

How Interamerican Greece stepped up its approach to model lifetime and manage value of its customers.



Dorte van Waarden
Analyst
Mlcompany



Peter Rozing
Program Manager
Mlcompany

achmea 

 INTERAMERICAN
PART OF ACHMEA

MI ϕ
COMPANY

Advanced lifetime modelling to value customers in a multiproduct environment

How Interamerican Greece stepped up its approach to model lifetime and manage value of its customers.

There is no industry in which value steering is so key as in the insurance industry. This is caused by the high cost of risk in this sector combined with the steep distribution between customers. However, value modelling at the customer level is very complex, driven by strong differences in customer behavior between and within business lines. As a result the majority of insurers currently steers on the margin per product. This means that differences in loyalty (lifetime) and the interaction effects between different products of a customer aren't adequately addressed in the valuation of customers. To overcome this, an overarching lifetime model is needed which gives accurate lifetime predictions on a customer level taking all information into account. In this article it is described which model is capable of creating such advanced lifetime predictions, which challenges may arise when building it and how a customer lifetime model can create value within organisations.

How customer lifetime modelling leverages value steering in organisations

Traditionally, most insurance companies steer their businesses on financial KPI's such as profit, sales and loss and cost ratios, which are monitored on a total portfolio or product level. However, these metrics only give high level insights in the company's current performance, whereas to optimize business decisions, companies need to understand the future value of their customers on a granular level.

Currently, more and more insurance companies are taking the leap into value-based business steering, in which having insight in the future value of an individual customer is key. Nowadays, this is ideally done by calculating the Customer Lifetime Value (CLV), a measure which quantifies the value of every single (potential) customer throughout the entire remaining customer lifetime. The power of using CLV in business steering is reflected in scientific research: this shows that businesses who steer on improving the total CLV of their customers (customer equity) also show positive financial results. In fact, a 10% increase in customer equity is amplified to a 15,5% increase in shareholder value (Schulze, Skiera & Wiesel 2012¹).

However, valuable as it may be, calculating CLV is not a job that is done easily. Since CLV is about future value, a combination of various advanced prediction models is required, instead of a relatively 'simple' calculation based on observed revenues, claims, costs et cetera. More specifically, in order to calculate CLV, an insurance company needs to know two things:

what is the total expected contribution margin for each customer for all products he or she has, and what is the expected remaining customer lifetime? Multiplying these numbers, and applying a discount rate to correct for the so-called 'time value of money', results in the CLV. Calculating the expected contribution margin per customer is something most insurance companies have learned to master over the past years. Actuarial departments have developed in-house capabilities to build advanced claim prediction models and cost allocation models, which, together with the premium, lead to the expected contribution margin per product per customer.

Using contribution margin for business steering is something that is done more and more by insurance companies. On the one hand, it allows them to focus their efforts on valuable customers, while on the other hand, it does not prevent them from maintaining the product orientation they are used to, since contribution margin is calculated per product. But this way of working calls for dramatic improvement: no attention is paid to the fact that the remaining customer lifetime greatly influences the total value of a customer. Moreover, the value of a customer arises from all products he or she is using and is not based on the contribution margin of one single product. The challenge is clear: in order to be able to perform advanced value steering, insurance companies should switch their focus from products to customers by modelling customer lifetime and calculating CLV.

This is easier said than done. Modelling customer lifetime is a task that comes with great

analytical challenges. However, if insurers manage to overcome these challenges, they will be able to leverage their value steering approach by focusing on the real value of their (potential) customers.

Building an advanced customer lifetime model

When facing the challenge of building a customer lifetime model, the course of action most people think of first is to build a churn model. Since customer lifetime depends on when a customer will churn, this seems to be the quantity to model, at least on first sight. But building such a model is, however, insufficient if you want to understand customer lifetime, since it focusses on churn within a specific time frame (mostly one year), while customer lifetime should be calculated based on the expected churn for multiple small time frames. There is no reason to assume that the expected churn per customer is constant over time. In fact, this typically is not the case: in most practical examples the probability that a customer leaves decreases over time.

So what should we do to model customer lifetime? The answer lies in survival analysis, where the aim is to predict the entire survival curve of a customer, based on the products he or she is using and a broad range of customer characteristics. The survival curve of a customer describes the probability that a customer is still a customer (he or she has 'survived') for each moment in time. Hence, this curve can be used to calculate the expected remaining lifetime of a customer by calculating the area underneath the graph. See figure 1 for an example of a survival curve.

Simple as this graph may look, modelling a survival curve is not an easy job to do, especially when attempting to create an advanced model

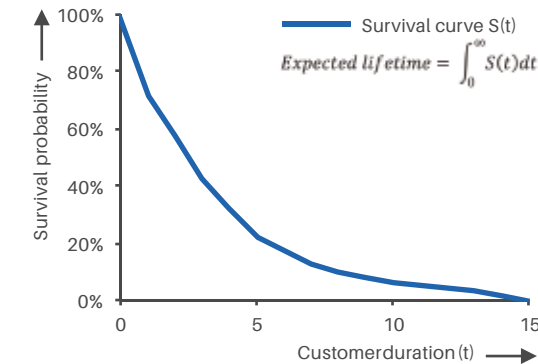


Figure 1.

by capturing all significant effects that are present in the data. This is, however, essential for creating a lifetime model that is suitable for business usage: simple models that only give good estimations for the entire population but fail for specific customer segments will eventually lead to unreliable predictions of the customer lifetime, which will hinder it from full adoption in and usage by the business.

Let's examine this in more detail. In general, there are two main challenges that may arise when predicting customer lifetime. The first one is due to the fact that customers aren't the same, but all are separate individuals who differ in various ways, such as product posses-

sion, social background, family composition and age, so there is no reason to assume that loyalty will be the same for every individual. Trying to model the survival probabilities for all customers by using only one model is often impossible and leads to quite accurate predictions for groups that are overrepresented in the customer base, but also to large deviations for groups that are more uncommon. This is undesired since, to be able to use the model outcomes for value creation, accurate predictions are required for all customers.

The second challenge relates to the nature of survival analysis: the aim is not to predict the churn at one certain moment, but to model the entire survival curve and hence the related survival probabilities for all moments in time. As was illustrated in figure 1, the theoretical (expected) survival curve is typically a smooth curve, but this does not necessarily match with actual survival rates.

Let's consider the example where the churn in the first years after inflow is very high (20% to 30%). However, after a couple of years the remaining customers tend to be very loyal and the churn rate decreases to only 5%. Because of the smooth shape of the survival curve it is not possible to estimate this relatively 'sharp angle' between the first three and the later years. This is illustrated in figure 2, where we see the estimated and realised outflow per customer duration. We see that new customers had an outflow percentage of 30%, whereas the model predicted only 8%. So even the model with the best fit was not capable to accurately estimate the outflow of customers who were active for less

than three years. However, the predictions for customers with a duration of more than three years turned out to be quite accurate. If such a model would be implemented this would result in accurate estimations of the remaining lifetime for existing customers who are active for more than three years, but also in a big overestimation of the lifetime for new customers. This would off course lead to misleading insights and wrong conclusions or actions.

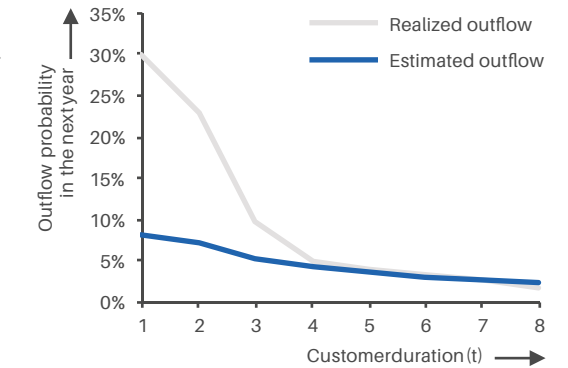


Figure 2. Realized and expected outflow in the next year, per customer duration - one model (best fit)

So how to overcome these problems? Which techniques can be used to create a survival model that is capable of capturing differences between customer groups and of capturing survival differences through time? The answer lies in grouping the customers together in the right way and in combining different models instead of trying to estimate all survival curves with only one model.

Capturing differences between customer groups

One of the characteristics of survival analysis is that it allows for grouping customers together, which is called stratification. Based on their characteristics, customers are divided into groups (strata), where customers within one stratum are expected to be relatively more similar to each other than to randomly chosen customers from other strata. In the insurance industry these strata are often based on the type of insurance that a customer has, where in other industries the age of a customer might be the most distinctive feature.

The benefit of stratification is that a stratified model is able to take loyalty differences between customer groups into account. For each stratum, a base survival curve is constructed. This curve reflects the average survival rate through time for customers within a stratum, thereby allowing these curves to differ over the various strata. Customers within a stratum have their own survival curve based on specific characteristics (age, social class, cover, etc.), but this survival curve can only deviate from the base survival of their stratum in a limited way. This shows that it is extremely important to perform stratification correctly; when customers who differ in terms of loyalty are assigned to the same stratum, the base survival curve is the same and the only differences in loyalty between these customers can be approached by small deviations from this baseline. Mostly, these deviations are too small to account for actual loyalty differences, and the survival model will lead to inaccurate predictions. Once the stratification is performed in the correct

way, the next step is to construct the correct model per stratum. One way to do this is to use the same model for all strata. This means that the same explanatory variables are included for all customers, stating that all customers have similar drivers for survival and only the exact weight of the drivers can differ. In practice, this often doesn't hold, as it can be the case that the age of the customer is the most important predictor of survival for customers with only a car insurance, where for Life insurance policies age is not significant at all, but the agreed term of the policy is.

Using one model often results in reasonable predictions for the groups (strata) with the most volume, but with large deviations for the smaller segments. This means that you could have a good prediction for all customers with product A (which is 80% of your base), but large deviations for the other customers. This challenge can be overcome by building a separate model for each stratum, but this may result in a proliferation of models, so it is important to cluster all strata into groups of similar explanatory variables. In this way, differences between strata can be taken into account in order to have good lifetime predictions for all customers.

Capturing survival differences through time

Since the survival curve is a smooth curve, it is practically impossible to end up with an appropriate prediction when the churn behavior in the first years after inflow is significantly different from the behavior in the later years, as was shown in Figure 2a. This example shows that it can be necessary to use a combination

of models within one stratum to obtain the best prediction. In this example case one model was built to predict the outflow percentages in the first three years after inflow. Another model was created that predicted the outflow from year four and onwards. Subsequently, both models were 'glued' together to find the survival probabilities between year three and year four. The results of this approach are illustrated in Figure 3, showing that the model fit in the later years is unchanged, but that the accuracy of the prediction in the first three years has significantly improved.

Of course, there are more challenges that should be overcome before being able to build an advanced customer lifetime model. Firstly, an integrated customer database should be built which is necessary to create a longitudinal customer view. With such a view it is possible

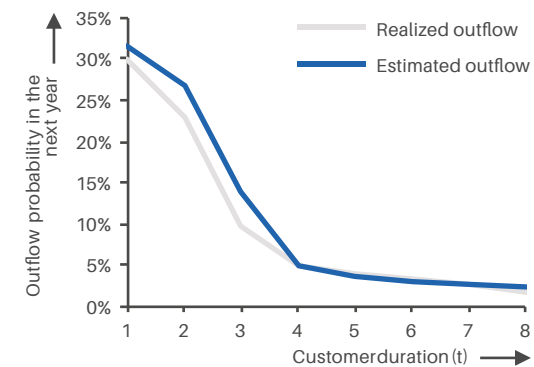


Figure 3. Realized and expected outflow in the next year, per customer duration - combination of models

to track the behavior of individual customers over time and to know which changes have occurred during his or her customer journey (for example changes in product possession or family composition). Having this information is crucial to construct accurate lifetime predictions. However, the task of creating a longitudinal customer view comes with some pitfalls. For instance, data warehouses are traditionally designed around products or business lines, and effort should be put into creating an integrated customer view, considering all products a customer is using. Moreover, this database should contain as much relevant (historical) information as possible, in order to be able to find the best survival curves and the most distinctive strata. Examples of information that has proven to improve survival models are customer variables like social class, educational level, and in the case of B2B-business number of employees, but also information about business and market changes.

Secondly, building a customer lifetime model comes with the risk of creating a black-box model that is only understandable to analysts, suffers from unexplainable correlations, and is therefore not usable for commercial purposes. An example of this was found when a survival model showed that inflow product (the customer's first product) was a very significant predictor of his or her lifetime. Marketing could not explain this effect, so additional analyses were performed, which showed that product of inflow had an extremely high correlation with year of inflow: before a specific date new customers typically joined with Health and after that date with Property - caused by a deliberate

business decision to change the focus of marketing efforts. The actual effect on customer lifetime was caused by year of inflow, and the model had to be adjusted to make sure that it would not predict unrealistically high lifetimes for new customers joining with Health.

Creating value from customer lifetime models

The importance and the benefits of customer lifetime modelling were also recognized by insurance company Achmea and their Division International (DI), which manages a portfolio of five operating companies in different countries. From 2014 onwards DI made the strategic choice to focus on Big Data and fact-based marketing. A critical part in this process was the development of the Customer Lifetime Value and hence the modelling of the expected lifetime per customer. Initially, the operating companies struggled to obtain predictions with the required accuracy, but with the help of Mlcompany, their survival predictions were adjusted using different models per strata and also different models within strata. This resulted in significant improvements in the predictions of survival, thereby revealing various opportunities for value creation.

As illustrated earlier, the expected customer lifetime is a crucial aspect of the CLV. With this model in place it is possible to understand exactly how each customer contributes to the future value of the company. Various analytical deepdives can be performed to uncover the performance of the different customers and to understand where value is created and destroyed. These insights are the foundation of various impact programs which help the OpCo's

to address the current business challenges, as is illustrated by three examples.

Targeted acquisition

As is the case for most companies, acquisition is one of the main drivers of business growth. For the Greek OpCo of DI, Interamerican Greece, acquiring new customers is even more essential since they have the very ambitious growth target to double the number of customers that is acquired through one specific sales channel. While the ambition is to increase the number of customers, at the same time this should not affect the customer equity (total CLV of the company). To balance both the value and the size of the portfolio it is essential to understand how the value of new customers differs, for example per age, region and product. A first decile analysis showed that the total value created by 80% of the new customers was destroyed by the bottom 20%, meaning that if those 20% with the lowest value hadn't been acquired the portfolio growth would be realized with a neutral effect on the total customer value. This 20% was characterized by both a negative margin and a high expected lifetime, resulting in a very negative CLV, whereas the top 80% was a combination of customers with a positive margin (and different lifetimes) and customers with a negative margin who were expected to stay relatively short. Based on these insights, the differences between the bottom 20% and the other 80% of the new customers were analysed, resulting in a detailed view of the profile of valuable prospects who should be targeted for new acquisition (and who not to approach). By shifting the acquisition focus to these valuable segments and not approach-

ing prospects with an expected very negative value, the ambitious growth targets can be met, while it is ensured that no customer equity will be destroyed.

Cross-sell and product bundling

One of the main insights from the survival predictions was that the bundling of insurance products was a crucial factor for increasing the loyalty. For example, customers with only a Home or only Earthquake insurance had an average lifetime of two years. However, when those products were combined, the loyalty increased significantly, resulting in an average customer duration of six years. This underlined the importance of cross-sell, not only to increase the margin, but mostly to commit the customer to the company and build a sustainable relation. Based on this insight the marketing intelligence team is currently focussing on creating a broad cross-sell impact program, by identifying the customer segments that are most likely to respond to the cross-sell offers and for which this has a positive effect on both margin and lifetime. Approaching and converting these customers will result in an increase in their CLV.

Churn reduction

Retention is one of the main challenges the OpCo's are facing. It is a serious problem in countries facing low insurance penetration and, with an overall churn percentage of 30% per year, this significantly influences the value of the portfolio. However, in the process of policy cancellation 60% of all customers gets in contact with the call center, meaning that there is a direct possibility to retain the customer by

offering the right benefits, discounts or a more appropriate insurance product. To make this retention process as effective as possible it is crucial to have a focused team of dedicated call center agents who have received proper training, have the right customer insights, tooling and know how to develop the customer journey further (so-called 'scripts'). For the development of the scripts it is essential to know which offer should be made to which customer, taking into account that the CLV of this customer will be positive, but also considering which offer will benefit the customer and ensure they are retained. In practice this often means that negative valued customers won't receive an offer at all, because this is never profitable. On the contrary, customers with a positive margin and a high expected lifetime can receive a substantial discount or benefit, for example in the form of an additional cover. These examples all show the value of steering on a forward looking customer level instead of on a backward looking product level. This approach is new for the OpCo's, but already proven to be very impactful, which is showed by the fact that the first sizings of the various opportunities resulted in a future potential ranging from €20 to €28 million per OpCo. This impact is also acknowledged by Suzanne Akten, Senior Manager – Marketing, Commercial and Customer Centricity of DI:

"We needed to accelerate profitable growth and it was clear following a traditional product driven view on the business would not align with our ambition of placing the customer centrally in our OpCo's and steering on future value. Together with Mlcompany we have started the journey to

implement and lead the businesses on Customer Lifetime Value. This process has challenges, but the insights created have had a significant impact, not only on the way we work but also how we review performance. Building a holistic forward looking customer view on the portfolio has given us insights into lifecycle behavior and highlighted the commercial shortfalls, which were previously not evident. This model spotlights how important it is to steer the business with a vertical strategy across all business lines – responding to customer behavior and creating logical customer journeys that encourage CLV growth. This journey has shown the importance of balancing portfolio value and volume. Applied appropriately, it can create stealth movements in the market that are hard to replicate and create a competitive edge of between 3-5 years in our local OpCo's."

Constructing an advanced customer lifetime model can be quite a challenge, but with the advanced modelling skills, an integrated customer view and a structured approach this is well within reach. With an accurate lifetime prediction it is possible to create a clear view on the value of all customers and hence to create a perspective on opportunities. Considering the value potential of these opportunities, it seems inevitable that companies will switch to this holistic way of doing their marketing. The only remaining question is who will lead and who will follow.

Appendix: Theoretical background of survival analysis

Survival analysis has the aim to predict the time until the occurrence of a certain event of interest. This event can be death (which

explains the name), but also marriage, cross-sell or churn. For simplicity we will use the terminology of customers and churn for the rest of the theoretic explanation. Assume that T is a continuous random variable with probability density function $f(t)$ and cumulative distribution function $F(t)=P(T<t)$, hence stating the probability that the customer has churned before time moment t .

Then, the probability that a customer who is active for a time t leaves in the short interval dt after t is equal to $P(t \leq T < t+dt | T \geq t)$. However, our interest is focused on the instantaneous rate of leaving, per unit time period at t . That is where the so-called hazard function falls into place, which is defined as

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

Globally stated, the hazard function is the probability that the customers churn in a certain period, given what has happened before that period. We note, however, that the hazard function is not a true probability in the sense that it can exceed the value 1 when Δt decreases.

By using the definition of conditional probability, we can express the hazard function in terms of the distribution and probability density function of T :

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t) \frac{1}{P(T \geq t)}}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{1 - F(t)} \frac{1}{\Delta t} = \frac{F'(t)}{1 - F(t)} = \frac{f(t)}{S(t)}$$

Where $S(t) = 1 - F(t) = P(T > t)$ is called the survival function, since it gives the probability of

survival to time t . If we are interested in the total expected customer duration, we than have by definition:

$$E(T) = \int_0^{\infty} t f(t) dt = \int_0^{\infty} S(t) dt$$

Where we use integrating by parts, and the fact that $f(t) = -\frac{dS(t)}{dt}$, which has limits $S(0)=1$ and $S(\infty)=0$.

Cox proportional hazards

Up to this point we have considered a homogeneous population, where the lifetimes of all customers are illustrated by the same survival function $S(t)$ and hazard $\lambda(t)$. However, customers have distinctive features, such as age, gender and product possession, which are likely to influence their lifetime. To cope with this the semi-parametric Proportional Hazards model is used, which is developed by David Cox (1987²). This method is viewed as empirically so successful that it has become the standard method for analysing survival data³. In the proportional hazards model (PH) the hazard at time t for an individual with covariates x_i (not including a constant) is assumed to be

$$\lambda(t|x_i) = \lambda_0 \varphi(x_i)$$

Hence, the effect of time is separated from the effect of the covariates. The time-dependent function $\lambda_0(t)$ is the baseline hazard function that describes the risk for individual with $x_i=0$. The function $\varphi(x_i)$ is the relative risk associated with the set of characteristics x_i , which is the same at all durations t . Usually $\varphi(x_i)$ is chosen to be equal to $e^{\beta^T x_i}$, since it ensures as positive hazard. Furthermore, it permits coefficients to be easily interpretable: suppose that the j th regressor x_j increases by one unit, while

the other regressors remain unchanged. Then:

$$\lambda(t|x_{new}) = \lambda_0(t) e^{\beta^T x + \beta_j} = e^{\beta_j} \lambda(t|x)$$

Thus, the new hazard is e^{β_j} times the original hazard.

Strata

One of the benefits of survival modelling is that it allows for stratification. This means that subjects (customers) can be divided into strata, where subjects within a stratum are expected to be relatively more similar to each other than to randomly chosen subjects from other strata. The regression parameters are assumed to be the same across the strata, but a different baseline hazard may exist for each stratum. In the insurance branch customers are often classified based on their product combination.

Sources

1. C. Schulze, B. Skiera & T. Wiesel, "Linking Customer and Financial Metrics to Shareholder Value: The Leverage Effect in Customer-Based Valuation", Journal of Marketing, Volume 76 (March), 2012.
2. D.R. Cox and D. Oakes, "Analysis of Survival Data", Chapman and Hall, London, New York, 1984.
3. A. Cameron and P. Trivedi, "Microeconomics; methods and applications", Cambridge University Press, 2005.
4. A. Tinazzi, M. Scott, A. Compagnoni, "A gentle introduction to survival analysis", PhUSE 2008, Paper ST03.