Work Project presented as part of the requirements for the Award of a Master Degree

in Finance from the NOVA – School of Business and Economics

CUSTOMER LIFETIME VALUE (CLV) MODELING

IN RETAIL BANKING

TOMÁS DE ALMEIDA DOS SANTOS (nr. 3278)

Project carried out on the Master in Finance Program, under the supervision of:

Professor Gonçalo Rocha

3rd of January, 2018

Abstract

Based on regression models, simple customer's attributes (age, income, assets and debt) - which banks usually use to identify who their most valuable customers are - were found not to be very effective at explaining and predicting customer's Gross Income. Thus, banks are recommended to consider alternative methods. A CLV estimation model based on Markov Chains is presented and tested as a potential alternative, even though our application is still rather conceptual, with limitations which would have to be addressed in future research. Also, another methodology based on retention cohort analysis is presented, aimed at estimating CLV for individual products.

Keywords: Customer Lifetime Value, Retail Banking, Markov Chains, Retention Cohort Analysis

1- Introduction and Business Challenge

This Work Project aims at developing Customer Lifetime Value (CLV) estimation methodologies appropriate for application in firms selling subscription-based products/services, with particular focus on retail banking. CLV of a customer may be defined, in simple terms, as the discounted value of the future cash-flows a firm is expected to capture from that customer.

Our partner bank was interested in CLV estimation for two distinct applications, which we refer to as *Application A* and *Application B*. *Application A* is the CLV estimation for individual products, answering to the question "when a customer subscribes product *X*, what is the expected value to be captured from this subscription?". This CLV measure may be later used by the bank to decide on acquisition and retention investments (such as offering gifts and vouchers to customers). *Application B* considers the complete relationship of the customers with the bank (considers all the products), answering to the question "what is the expected value to be captured from customer *X* in the next years?". Having a more accurate measure of customers' value would allow the bank to take more informed Marketing decisions, such as deciding which customers are worth assigning an account manager.

We start by presenting a *Retention Model* for *Application A*, which, as explained in Literature Review, is appropriate to model cases in which there is a binary concept of retention (the customer is either retained or not in each period), and, when the customer churns in the product, it is considered that he or she will no longer generate any value.

Concerning *Application B*, we start by testing how effective is the simplest method that banks usually use to identify their most valuable customers: by considering customer's income and financial assets. Thus, we estimate an *Explanatory Regression Models* to test whether these and other two simple variables (customer's Age and Debt) are indeed good explainers of customer's Gross Income, in a given year. We also estimate a *Predictive Regression Model* to test whether these variables are good predictors of next year's Gross Income. If these hypotheses are validated, there is a case to be made that it is not worth investing resources to develop a more sophisticated CLV model.

Finally, we develop a *Migration Model* based on *Markov Chains* to estimate CLV, which may be the basis for an alternative model to be used by the bank to identify their most valuable customers. A Migration Model, as described in Literature Review, is appropriate to model cases of firms with *always-a-share* relationships with their customers, in which there is always a probability that a customer generates some value in the next period (even if very low). A Migration Model allows for the fact that, each period, a customer may increase or decrease his/her *engagement level* to a different level, for example, by upgrading/downgrading a subscription, or by changing the level of usage. Due to some limitation which are later explained, the results of the applications are not yet reliable, so the contribution of this model to solve the bank's challenge is more conceptual and exploratory in nature.

2- Literature Review

Pfeifer et. al (2004) dedicated a paper to the clarification of the inconsistencies when authors refer to two related but different concepts: *customer lifetime value* and *customer profitability*.

Customer Lifetime Value (CLV) is closer to the concepts of *Present Value* and *Valuation* from Finance Theory. It is the discounted value of the future expected cash-flows a firm is expected to capture from a customer. Thus, CLV has a forward-looking approach, as it captures the expectation of evolution of the relationships customers and time value of money. On the other hand, *customer profitability* is closer to the concept of *Accounting Profit* from Financial Accounting. It is the Gross Income the customer historically generated to the company, during some period in the past. Just as in the financial valuation of a company, a customer may currently be unprofitable, but still be valuable (**Pfeifer et. al,** 2004).

The emergence of the CLV concept is linked to the emergence of Customer Relationship Management (CRM), the subject concerned with driving value from relationships and loyalty (**Haenlein et al.**, 2007), by focusing on the relationships with customers globally, instead of focusing on single isolated transactions. There has been a transition from a *product-centric view*, in which products are the key firms' assets and the focus is on selling them, to a *customer-centric view*, in which customers are the key assets and the focus is on retaining and capturing more value from them (**Jain & Sign**, 2002).

A key principle in CRM is to treat customers differently, depending on their potential value (**Haenlein et al**., 2007; **Malthouse & Blattberg**, 2005). To be able to take decisions on resource allocation to customers, the first challenge is to have an appropriate measure of the value of each customer: CLV. Some of the industries in which companies may leverage CLV to take better decisions are: airlines, retail stores with loyalty programs, internet services, telecommunications, catalogue sales, media publishing, software (**Ekinci** et al, 2012).

2.1- Mathematical modeling

An important starting point when modeling CLV is to define which type of buyer-seller relationship best describes the firm in analysis (**Dwyer**, 1989 and 1997). The author cited **Jackson's** (1985) simplified dichotomy in which there are only two types of buyer-seller

relationships. In a *lost-for-good setting*, when a customer leaves the firm, he/she has taken a long-term commitment and will never return back. It is appropriate for a contractual or subscription products/services. In an *always a share setting*, customers allocate part of their total consumption in a product category to different providers, without contractual commitment to any particular one. No purchases from a customer in a period does not imply that the customer will never purchase again in the future periods (such as in retail industry).

As an alternative classification, **Reinartz & Kumar** (2000) define two types of relationship setting: a *contractual setting*, when there is a commitment and higher predictability; and a *non-contractual setting*, when the customer may split his/her expenditures to different providers and easily switch. As stated by **Fader & Hardie** (2009), in contractual settings the customers' churn is observed and verifiable in each point in time. In a non-contractual setting, a customer silently churns, so it is not possible to accurately verify how many customers the firm has. Instead, there are proxies such as "customer who purchased last month".

According to **Dwyer** (1989 and 1997), each of these two types of buyer-seller relationship require a different type of CLV model. Lost-for-good and contractual settings are best modeled with a *Retention Model*, in which the customer may either be retained or not, in each period. Expected retention rates may be found empirically using cohort analysis of customers who became customers in previous periods. Always-a-share and non-contractual setting are best modeled with a *Migration Model*, in which there is always a probability of the customer buying in the next period (even if very low), for example, depending on the *recency* of the last purchase.

Berger & Nasr (1998) considered that the field of CLV was lacking well-grounded quantitative methods, as previous research was more focused on the qualitative discussion of the topic and only presented simple numerical examples. The authors gave a contribution to the field by developing six general mathematical models to estimate CLV. Each model is more suitable to a specific situation, described with a set of assumptions, for example, regarding the

timing of revenues, the growth pattern of revenue for retained customers, whether the cash-flows are discrete or continuous.

Regression models have been used to model CLV (**Matlhouse & Blattberg**, 2005). For example, **Ekinci** et al (2012) implements a regression model in the banking industry, to predict the profit per customer in the next year, using activity/usage variables and product ownership as explanatory variables.

In 2000, **Pfeifer & Carraway** introduced a new class of CLV models: Markov Chain Models, using a mathematical concept developed by **Markov** (1906) with successful academic application in different fields of study (**White**, 1993). In a Markov Chain, there is a set of possible states (the Markov states) and, in each period, there is a probability associated to transitioning from the current state to each of the possible states. These are called *transition probabilities* and may be presented in a *transition matrix*. One of the examples consider that the Markov states are customer's Recency of last purchase (last purchase in t-1, t-2,...). Given a customer's initial state, there is an estimation of his/her CLV, which is the result of the expected value of purchasing or not in each of the following periods.

Another application of Markov Chain is found in **Rust et. al (2004)** research. In this case, the authors considered the Markov states to be the available brands in the market. The transition probabilities are the probabilities that customers switch between brands from a period to the other. **Haenlein et al. (2007)** also applied Markov Chains to model CLV, in the context of retail banking. The authors considered the Markov states to be customer segments. There is a CLV estimate for each initial Markov state (customer segment).

3- Methodology

3.1- Retention model with cohort analysis

As CLV depends on uncertain future cash-flows, when authors commonly refer to "CLV" they are actually referring to the expected value of a random variable CLV (Pfeifer et. al, 2004), which is a linear combination of other random variables, namely the random variables CF_t - the

cash-flow the firm captures from the customer in period *t* - properly discounted to the present. Expression 1 is the most general CLV model and it is an important starting and reference point, as more complex models are particular cases of this general formulation, aimed at answering how the random variables CF_t are estimated.

$$\begin{aligned} CLV_0 &= PV(CF_1) + PV(CF_2) + \cdots + PV(CF_t) = \sum_{i=1}^{t} PV(CF_t) = \\ &= \frac{1}{(1+r)^1} CF_1 + \frac{1}{(1+r)^2} CF_2 + \cdots + \frac{1}{(1+r)^t} CF_t \\ &= \sum_{i=1}^{t} \frac{1}{(1+r)^t} CF_t \end{aligned} \tag{1}$$

$$E(CLV_0) = PV[E(CF_1)] + PV[E(CF_2)] + \cdots + PV[E(CF_t)] = \sum_{i=1}^{t} PV[E(CF_t)] \tag{2}$$

A CLV model must account for time value of money and risk with an appropriate discount rate, as future cash-flows are uncertain and risky. For simplification purposes, we account for time value of money, throughout the Methodology section, by multiplying cash-flows by a conceptual discount factor, PV(), equivalent to $\frac{1}{(1+r)^t}$.

The general format of a *Retention Model* is presented in Expression 3. The random variable *Cash-flow* (CF_t) for a period is the result of multiplying the random variable *Retention Rate* (R_t) of the same period by the product's Gross Income.

$$\begin{aligned} CLV &= PV(CF_1) + PV(CF_2) + \cdots + PV(CF_t) \\ &= PV(GC * R_1) + PV(GC * R_2) + \cdots + PV(GC * R_t) \\ &= PV(GC * \sum_{1}^{t} R_i) \end{aligned} \tag{3}$$

A *Cohort Analysis* is a method commonly used by practitioners to empirically estimate the random variables *Retention Rate* (R_t), for each period. In CLV academic research, this method was mentioned by Dwyer (1997). The first step is to define *cohorts*, for example, cohorts formed by the customers who subscribed the service in each month. Then, for each cohort, it is observed how many customers are retained in each of the following months. The number of

customer still retained in a given month divided by the initial number of customers in the cohort is the retention rate for that month (for that cohort).

				Months after subcription			
Cohort	**Observations in cohort**	**Relative frequency**	**0**	**1**	**2**	**…**	**t**
A	N_A	$f_A = N_A/N$	100%	R_1^A	R_2^A	…	R_t^A
…	…	**…**	…	…	…	…	…
n	N_n	f_n	100%	R_1^n	R_2^n	…	R_t^n
Total and Weighted average	$\boldsymbol{N}$	100%	100%	$E(R_1)$	$E(R_2)$	…	$E(R_t)$

Summarizing the information in a table with the same format as the table above, gives the firm visibility into the evolution of retention overtime and allows to estimate the expected value and variance of the random variables *Retention Rate* (R_t), for each period. Cohorts are labeled from A to *n*. R_t^A is the observation of retention rate in the period *t*, for the first cohort (A). f_A is the relative frequency of the cohort: the number of customers in the cohort divided by the total number of customers (all cohorts) that existed up to that period. Expression 4 is used to estimate the expected value and Expression 5 to estimate variance, for the random variables.

$$E(R_t) = [f_A \quad f_B \quad \dots \quad f_n] \begin{bmatrix} R_t^A \\ \vdots \\ R_t^n \end{bmatrix} \tag{4}$$

$$Var(R_t) = [\begin{bmatrix} f_A & 0 & \dots & 0 & 0 \\ 0 & f_B & \vdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \vdots & \ddots & 0 \\ 0 & 0 & \dots & 0 & f_n \end{bmatrix} \begin{bmatrix} R_t^A - E(R_t) \\ \vdots \\ R_t^n - E(R_t) \end{bmatrix}]^T \begin{bmatrix} R_t^A - E(R_t) \\ \vdots \\ R_t^n - E(R_t) \end{bmatrix} \tag{5}$$

The random variable Cash-flow (CF_t) for a period is obtained by multiplying the random variable *Retention Rate* of that period by the products' Gross Income (Expression 6).

$$CF_t = GC * R_t \quad (6) \quad ; \qquad \sigma_{CF_t} = GC * \sigma_{R_t} \quad (7)$$

The final step is the estimation of the random variable *CLV*, which is the addition of the random variables *Cash-flow* (CF_t) of all the *t* periods considered (Expression 8).

$$CLV = PV(CF_1) + PV(CF_2) + \cdots + PV(CF_t) = PV(\sum_{i=1}^{t} CF_t) \tag{8}$$

The standard deviation of the random variable *CLV* is estimated with Expression 9. It depends on multiple covariances between the random variables *Retention Rate* (R_t). w is a vector of 1's, with the number of rows equal to the number of periods considered, so with dimensions 1*t and $\sum$ (Expression 10) is the covariance matrix between the random variables *Retention Rate* (R_t). Expression 12 is an example to present how the covariances are computed.

$$\sigma_{CLV} = \sqrt{w^T \sum w} \quad (9) \quad ; \quad \sum = \begin{bmatrix} Covar(R_1, R_1) & \cdots & Covar(R_1, R_t) \\ \vdots & \ddots & \vdots \\ Covar(R_t, R_1) & \cdots & Covar(R_t, R_t) \end{bmatrix} \quad (10)$$

$$Covar(R_1, R_t) = \left[\begin{bmatrix} f_A & 0 & \dots & 0 & 0 \\ 0 & f_B & \vdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \vdots & \ddots & 0 \\ 0 & 0 & \dots & 0 & f_n \end{bmatrix} \begin{bmatrix} R_1^A - E(R_1) \\ \vdots \\ R_1^n - E(R_1) \end{bmatrix} \right]^T \begin{bmatrix} R_t^A - E(R_t) \\ \vdots \\ R_t^n - E(R_t) \end{bmatrix} \quad (11)$$

More granularity may be obtained by estimating CLV for different customer segments, based on their characteristics and behavior. For example, by aggregating customers based on demographic attributes, purchase channel, product usage, ownership of other products.

3.2- Regression models

As explained in Introduction, we estimate regression models to test two hypotheses. First, whether customer's *Age*, *Income*, *Assets* and *Debt* variables are indeed good explainers of customer's Gross Income in a year. Second, whether these variables are good predictors of next year's customer's Gross Income.

The first hypothesis is tested with an *Explanatory Regression Model* (Expression 12), having *Gross Income* as dependent variable and *Age*, *Income*, *Assets* and *Debt* as explanatory variables, for the same year. *Age* is the age at the end of the period and *Income*, *Assets* and *Debt* are the averages during the year. The second hypothesis is tested with *Predictive Regression Model* (Expression 13), to predict the next year's customers' Gross Income, with the same independent variables. Customer's Gross Income ("Produto Bancário", in Portuguese) is the accumulated amount for the year and is a result of the sum of net financial income (driven by the customer's Assets and Debt) and net non-financial income (other fees and commissions). In order to find

the most suitable model, significant interaction variables (between the four variables) are also included in the models.

Explanatory regression models (same year): $Gross\ Income = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Income} + \beta_3 \text{Assets} + \beta_4 \text{Debt} + \varepsilon$ (12)

Predictive regression models (one-step ahead): $Gross\ Income_{t+1} = \beta_0 + \beta_1 Age_t + \beta_2 Income_t + \beta_3 Assets_t + \beta_4 Debt_t + \varepsilon$ (13)

3.3- Migration model with Markov Chains

Now we present a model which may be used by the bank as a basis to develop an improved alternative method of estimating CLV. This methodology is based on the research by Haenlein et al.(2007), mentioned in Literature Review, which uses the mathematical concept of Markov Chains to model the customers' relationships with the firm.

A *Markov Chain* is a process with a discrete set of possible *Markov states* ($\{s_1, \dots, s_n\}$) which starts in one of the states, and, in each of the following period, may move to any of the states. The probabilities of transition between states are called *transition probabilities* and are usually presented in a matrix (a *transition matrix*) with dimensions *n*n*, being *n* the number of Markov states. If the transition probabilities are only conditional on the previous state, it is called a *first-order* Markov Chain (memoryless). Expression 14 is the general format of a transition matrix, in which, for example, $P(s^{t=1} = s_2 | s^{t=0} = s_1)$ is the conditional probability of being at state 2 in the next period, given that the previous state is 1.

$$P_1 = \begin{bmatrix} P(s^{t=1} = s_1 | s^{t=0} = s_1) & \cdots & P(s^{t=1} = s_n | s^{t=0} = s_1) \\ \vdots & \ddots & \vdots \\ P(s^{t=1} = s_1 | s^{t=0} = s_n) & \cdots & P(s^{t=1} = s_n | s^{t=0} = s_n) \end{bmatrix} \quad (14)$$

Markov Chains have an application to CLV modeling, if the evolution of the customers' relationships with the firm are viewed as Markov Chains. There are *n* possible *engagement levels* in which customers may be in each period (the Markov states). In the beginning of the relationship with the firm, the customer is at one of these states, and then, in each of the following period, the customers may move to other states (including terminating the

relationship). Instead of estimating CLV for each individual customer, this method allows to estimate CLV for each of the *n* states, in which customers are allocated to.

A probability tree is a useful visual way for understanding how a Markov Chain evolves. Appendix 1 is an example with only two states and two periods. A customer's relationship with the firm may take multiple "paths" overtime, which grow exponentially. For example, for a customer to be at state 1 in the first period, there are two possible "paths" which result in that state, so the total probability $P(s^{t=1} = s_1 | s^{t=0} = s_1)$ is given by the sum of two probabilities. Then, there are 4 possible paths for a customer to be at state 1 in the second period.

[Appendix 1]

In simple terms, the CLV of a customer is the expected value of the Cash-flows the customer generates, capturing the multiple possible "paths" the relationship may take. Thus, we want to estimate the probabilities for all those paths. The initial *transition matrix* presents the probabilities of a customer being at each state, in the first period, given any initial state. For each of the following periods, we want to calculate the total probabilities of the customer being at each possible state in that period, given any initial state. We calculate these total probabilities for the following periods by successively multiplying the initial transition probability by itself.

Expression 15 presents the general expression to calculate the *total probabilities matrixes*, for each period, and Expression 16 is an example of the total probabilities for the second period, for a simple case with only two Markov states. $p_{2,1}$ is a shorter way of representing $(s^{t=1} = s_1 | s^{t=0} = s_2)$.

$$P_t = {P_1}^t = \begin{bmatrix} P(s^{t=t} = s_1 | s^{t=0} = s_1) & \cdots & P(s^{t=t} = s_n | s^{t=0} = s_1) \\ \vdots & \ddots & \vdots \\ P(s^{t=t} = s_1 | s^{t=0} = s_n) & \cdots & P(s^{t=t} = s_n | s^{t=0} = s_n) \end{bmatrix} \tag{15}$$

$$\begin{aligned} P_2 &= \begin{bmatrix} P(s^{t=2} = s_1 | s^{t=0} = s_1) & P(s^{t=2} = s_2 | s^{t=0} = s_1) \\ P(s^{t=2} = s_1 | s^{t=0} = s_2) & P(s^{t=2} = s_2 | s^{t=0} = s_2) \end{bmatrix} = P_1^2 = \\ &= \begin{bmatrix} P(s^{t=1} = s_1 | s^{t=0} = s_1) & P(s^{t=1} = s_2 | s^{t=0} = s_1) \\ P(s^{t=1} = s_2 | s^{t=0} = s_1) & P(s^{t=1} = s_2 | s^{t=0} = s_2) \end{bmatrix}^2 \\ &= \begin{bmatrix} p_{1,1} * p_{1,1} + p_{1,2} * p_{2,1} & p_{1,1} * p_{1,2} + p_{1,2} * p_{2,2} \\ p_{2,1} * p_{1,1} + p_{2,2} * p_{2,1} & p_{2,1} * p_{1,2} + p_{2,2} * p_{2,2} \end{bmatrix} \end{aligned} \tag{16}$$

Having a total probabilities matrix for each period, we associate these probabilities to the monetary values of the respective states to calculate the expected cash-flows generated in each period, for each initial state. The matrixes CF_t represent these cash-flows and are obtained, for each period, by multiplying the total probabilities matrix for the period by a vector with the Gross Income associated to each state. Expression 17 is the example of the Cash-flow matrix for the second period.

$$\begin{aligned} CF_2 = P_2 * GI &= \begin{bmatrix} P(s^{t=2} = s_1 | s^{t=0} = s_1) & P(s^{t=2} = s_2 | s^{t=0} = s_1) \\ P(s^{t=2} = s_2 | s^{t=0} = s_1) & P(s^{t=2} = s_2 | s^{t=0} = s_2) \end{bmatrix} \begin{bmatrix} GI_1 \\ GI_2 \end{bmatrix} \\ &= \begin{bmatrix} P(s^{t=2} = s_1 | s^{t=0} = s_1) GC_1 + P(s^{t=2} = s_2 | s^{t=0} = s_1) GI_2 \\ P(s^{t=2} = s_2 | s^{t=0} = s_1) GC_1 + P(s^{t=2} = s_2 | s^{t=0} = s_2) GI_2 \end{bmatrix} \end{aligned} \quad (17)$$

Finally, CLV is a matrix with one column and *n* rows, as there is a CLV associated to each initial Markov State (Expression 18). If a customer is at state *n* initially, we conclude that his/her CLV estimate is equal to the CLV estimate for that state *n*.

$$\begin{aligned} CLV = \; & PV(CF_1) + PV(CF_2) + \cdots + PV(CF_t) \\ & = PV(P_1 * GI) + PV(P_2 * GI) + \cdots + PV(P_t * GI) = \begin{bmatrix} CLV_1 \\ \vdots \\ CLV_n \end{bmatrix} \end{aligned} \quad (18)$$

To apply this model to a retail bank, we must begin by defining the possible *Markov states*. This process was found to be very challenging because retail banks have the particularity of offering many different products, so that customers may have one of multiple possible *engagement levels* at each time, given by the combinations of the products (e.g., Debit and Credit cards, Investment products, Insurance products, etc). The number of *engagement levels* increases exponentially with the number of products offered. For example, if the bank offers 20 different products, there are 1.048.576 (2^{20}) possible combinations of products.

Product ownership is not the only challenge. The pricing and revenue streams generated by those products are very customer-specific, depending on usage and volumes in products. For example, two customers with the same credit card may have very different number of transactions and transacted volumes.

These two reasons – the amount of possible engagement levels and non-standard pricing – implies that CLV estimation of the customers' complete relationship in retail banking is a very complex problem. We test a simplified application of the methodology, while being aware of the limited value of the results, so the contribution of this Migration Model has conceptual and exploratory nature.

Defining Markov states based on product ownership would not be adequate for this study, as its computational implementation would be too heavy and complex. Instead, we define the Markov states as simple customer segments, defined by the intersection of three customer's attributes: customer's *Age*, *Income*, and *Assets+Debt* (the sum of customer's balance in investment products and in credit products).

Eleven ranges for Income and Assets+Debt ranges are defined: level 1 for low levels of Income/Assets+Debt (less than 500€), and then, the other ten levels are the variables' deciles (up to the top percentile 99,9%). Eight ranges are defined for Age: 26-30, 31-35, 36-40, 41-45, 46-50, 51-55, 56-60, 61-65. Thus, there are 121 (11*11) customer segments for each of the eight age group, resulting in a total of 968 micro segments (121*8).

There is a *transition matrix* for each of the eight age groups, presenting the empirical probabilities of movement across the 121 segments (from 2015 to 2016). For example, the transition probability from a level x to level y are the number of customers who moved from level x to y divided by the total customers in level x initially. An additional level (zero) is included, for the case in which the customer churns, so the transition matrixes have dimensions 122*122. Additionally, each of the 968 segments (Markov states) has an associated value, which is the average of the Gross Income generated by the customers in that segment. It is assumed that a customer in segment x generates to the bank the Gross Income equal to the average of the segment, so it would important for have a small standard deviation of Gross Income within each segment.

We estimate CLV for a time horizon of 3 years and, to make the computational implementation easier, we consider the same transition matrix for the three years. For example, for a customer with initial age 39, the *transition matrix* used is always the one for the age group 36-40, even if the customer is older than 40 years old in the second year.

4- Data

4.1- Dataset A: used in the Retention Model

Dataset A is a subset of the database related to the management of the product in analysis, which contains information about product's subscription, the monthly charge of the subscription price and about the status of the contract. A description of the selected subset of variables is presented in the Table 1.

Variable name	Description
product_contract_code	Internal code to identify the contract
subscription_date	Date in which the customer subscribed the product
subscription_month	Month in which customer subscribed the product
contract_status	Categorical variable: "Active", "Irregular" or "Canceled"
status_date	Date in which the contract_status changed for the last time
cancelation_month	If state_of_contract is "Canceled", this is the month of status_date
months_until_ cancelation	Variable given by: subscription_month - cancelation_month
month_1(/2/3/4…)	These are dummy variables: 0 if the contract is canceled in the month and 1 if not. These variables are obtained by applying conditions based on the other variables, namely contract_status and months_until_cancelation.

Table 1- Dataset A variables and description

4.2- Dataset B: used in the Regression Models

The original dataset provided by the bank is simplified to include only 6 variables. For the year 2016, *Age, Gross Income, (Average monthly) Income*, *(Average) Assets, (Average) Debt*, and also the *Gross Income* in 2017 (January to end of October). Each customer is an observation in the dataset. For confidentiality reasons, we avoid disclosing summary statistics.

Only a subset of the dataset is considered, in order to have a more homogenous group and the solely consider to customers who the bank has more interested in applying the methodology to. The final dataset contains: active customers; customers not in default; customers between 26 and 66 years old in 2016; customers who received income in the bank, and whose monthly income was not below 500€.

Gross Income is highly impacted by the financial component, driven by customer's Assets and Debt. Debt may have a particularly negative distortion, because customers may have contracted a loan many years ago with a low interest rate for the current economic context, which makes the customer unprofitable now. To decrease this distorting effect, we remove customers with significant Debt balances (more than 15.000€), as to remove customers with home loans while keeping customers with small short-term credit.

In order to decrease outliers, some observations are dropped: with *Gross Income* in the top 99% percentile and bottom 1% percentile; with *Income, Assets and Debt* on the top 99% percentile. The final dataset has above 500.000 observations. Regarding the distribution of the variables (Appendix 2), the variable *Age* is approximately uniform and the other variables are highly positively skewed. In order to decrease skewness, the variables are log transformed (except for *Age*). A constant 100 is added to Gross Income and 0,1 to the other variables, not to have non-positive values.

[Appendix 2]

In what concerns the relationships between variables, a correlation matrix (Appendix 3) summarizes the correlations between all variables. The correlation between *Assets* and *Gross Income* in 2016 is around 0,48; between *Gross Income* and *Debt* is around 0,35; between *Gross Income* and *Income* is around 0,27; and the other correlations are weak.

[Appendix 3]

Appendix 4 presents scatter plots to visualize the relationship between the variables. In general, the relationship between any two variables is weak.

[Appendix 4]

4.3- Dataset C: used in the Migration Model

Dataset C considers a wider subset of customers and more variables. For the years 2015 and 2016: *Age, Gross Income, (Average monthly) Income*, *(Average) Assets, (Average) Debt*. The final dataset includes: active customers; customers not in default; customers between 26 and 66

years old in 2016. Customers in the top 99,9% percentile for Income, Assets+Debt and Gross Income in any year are removed, as well as in the bottom percentile 0,01% of Gross Income.

5- Results

5.1- Retention Model

Information contained in Dataset A is summarized table (Table 2), which, as described in Methodology section, is a cohort analysis which gives visibility into how retention evolves every month after the subscription month. Only 6 months of data is available.

		Months after subscription					
Cohort	**Relative frequency**	**0**	**1**	**2**	**3**	**4**	**5**
Jun.	11,9%	100%	98%	97%	96%	95 %	94,3%
Jul.	18,0%	100%	97,9%	97,2%	96,5%	95,7%	-
Aug.	27,6%	100%	98%	97,2%	96,7%	-	-
Sep.	13,7%	100%	98%	97,4%	-	-	-
Oct.	15,0%	100%	98,2%	-	-	-	-
Nov.	13,9%	100%	-	-	-	-	-
Total and Avg.	100%	100%	98%	97,2%	96,5%	95,4%	94,3%

Table 2- Results for the retention cohort analysis

As described in Methodology, *Retention Rate*, for each month, may be described as random variables (R_t), with a certain expected value and standard deviation (Table 3). Table 3 presents the estimates of Expected Value and Standard Deviation for the random variables *Retention Rate* (R_t).

	R_1	R_2	R_3	R_4	R_5
Expected value	98%	97,2%	96,5%	95,4%	94,3%
Standard deviation	0,10%	0,13%	0,24%	0,32%	-

Table 3- Statistics estimates for the random variables *Retention Rate*

Table 4 presents the estimates for the random variables *Cash-flow* (CF_t), given by the multiplication of the random variable *Retention Rate*, for the same period, by the product's Gross Income. For confidentiality reasons, the actual monthly Gross Income value is not disclosed, and assumed to be 5€.

	CF_1	CF_2	CF_3	CF_4	CF_5
Expected value	4,90 €	4,86 €	4,83 €	4,77 €	4,71 €
Standard deviation	0,01 €	0,01 €	0,01 €	0,02 €	- €
Discounted Expected Value	4,86€	4,78€	4,71€	4,62€	4,53€

Table 4- Statistics estimates for the random variables *Cash-flow*

Finally, having the random variables of *Cash-flow* for each period after subscription, *CLV* is a result of the addition of those variables (discounted)- Table 5. We decided to compute the CLV only for a 5-month horizon, as there are only 5 months of historical data. As we see, if we simply ignored customer churn, we would assume that the Gross Income for the first 5 months was 5*5€=25€. However, when accounting for customer churn, the expected value of CLV is 24,07€ instead (ignoring discounting).

Product's Gross Income	5€
Expected Value CLV 5 months	**24,07€**
Discounted Expected Value CLV 5 months	**23,51€**
Discounted CLV Standard deviation	0,16€

Table 5- Summary of the CLV estimation results

5.2- Regression Models

5.3.1- Explanatory regression model (same year)

The regression model with best fit includes all the variables individually and all possible interactions between the variables. Gross Income 2016 is the dependent variable and the summary output is presented in Appendix 5. The model is globally significant, as well as all variables individually and the R-squared 0,51, so only around 51% of the variation in Gross Income is explained by the model.

[Appendix 5]

Table 6 compares the summary statistics of the Actual values, Estimated values and Residuals. The distribution of the estimated values is significantly less skewed than the actual values (Appendix 6). Regarding *Regression Diagnostic*, the residuals have mean zero, but its distribution is more peaked than normal distribution, with excess kurtosis 1,7 (Appendix 6); the scatter plot of Residuals against Estimated values suggests some heteroskedasticity (Appendix 7). Thus, the model does not perfectly conform with linear model assumptions.

	Min	1Q	Median	Mean	3Q	Max	St. Dev
log(Actual Values)	4,36	4,78	4,97	5,08	5,28	6,78	0,42
Estimated Values(log)	3,25	4,86	5,09	5,08	5,29	6,073	0,29
Residuals	-1,44	-0,20	-0,04	0	0,15	2,49	0,33

Table 6- Summary statistics for the Actual values, Estimated values and Residuals

[Appendix 6 and 7]

5.3.3- Predictive Regression model (one-step ahead)

The regression model with best fit includes all the varaibles individually and all possible interactions between the variables. Gross Income 2017 is the dependent variable and the summary output is presented in Appendix 8. The model is globally significant, as well as all variables individually and the R-squared 0,39, so only around 31% of the variation in Gross Income is explained by the model.

[Appendix 8]

Table 7 compares the summary statistics of the Actual values, Estimated values and Residuals. The distribution of the estimated values is significantly less skewed than the actual values (Appendix 6). Regarding *Regression Diagnostic*, the residuals have mean zero, but its distribution is more peaked than normal distribution, with excess kurtosis 1,1 (Appendix 9); the scatter plot of Residuals against Estimated values suggests some heteroskedasticity (Appendix 10). Thus, the model does not perfectly conform with linear model assumptions.

	Min	1Q	Median	Mean	3Q	Max	St. Dev
Log(Actual Values)	4,12	4,79	4,99	5,11	5,3	6,95	0,47
Estimated Values(log)	3,74	4,89	5,08	5,11	5,33	6,34	0,29
Residuals	-1,72	-0,21	-0,03	0	0,16	2,22	0,36

Table 7- Summary statistics for the Actual values, Estimated values and Residuals

[Appendix 9 and 10]

5.3- Migration model with a Markov Chain

As mentioned in Methodology, 968 customer micro segments are defined, given by the intersection of 11 levels of *Income*, 11 levels of *Assets+Debt* and 8 *Age* groups. As an example, Appendix 11 summarizes the percentage of the customer base in each segment, for the age group 51-55, and Appendix 12 the percentage of total Gross Income generated by each segment. Appendix 13 summarizes the average Gross Income of each segment, having as reference (1,00) the total average Gross Income of the age group. For confidential reasons, we do not disclose the actual monetary values.

[Appendix 11, 12 and 13]

Appendix 14 presents of the distributions of Gross Income for three of the segments, as examples. Even when considering 968 segments, Gross Income has a high variance within each segment and is highly skewed. Thus, as explained in Methodology, the number of segments/states considered is too small. As a consequence, the CLV estimates will also have high variance and skewness, so not providing reliable results.

[Appendix 14]

We empirically found a *transition matrix* for each of the 8 age groups, presenting the empirical *transition probabilities* across the 121 segments, which are the relative frequencies of the movements (from end of 2015 to end of 2016). Appendix 15 is an example of a subset of the transition matrix for the age group 51-55.

Finally, the application of the methodology results in a CLV estimate for each of the 968 segments. As an example, Appendix 16 presents the final CLV estimates for each of the 121 segments of age group 51-55 years. The CLV estimates have as reference index (1,00) the average Gross Income of the age group. The values are not discounted, in order to easily compare the 3-year CLV with the current segments' Gross Income.

[Appendix 16]

6- Conclusions and Discussion of Future Research

Two CLV models were presented to be applied to two of the partner bank's challenges. To estimate CLV for individual products (*Application A),* a *Retention Model* (based on retention cohort analysis) is presented; and, to estimate CLV for the customers' complete relationship with the bank (*Application B*), a *Migration Model* (based on Markov Chains) is presented. By estimating two *Regression Models*, we diagnosed how effective the bank's current method to evaluate customers' value is.

The *Retention Model* is built on top of a retention *cohort* analysis, which is a tool for the bank to have visibility into customer retention overtime, in the product. The CLV estimates serve as references when deciding on acquisition and retention investments. The methodology

is also general enough to be applied to other industries offering subscription products/services. For future research, it would be interesting to have CLV estimates with for each individual customer, based on their specific characteristics and behavior, which would be possible with probabilistic classification models (outputting the probability of retention in the next *n* periods, for each customer).

Regarding *Application B*, we started by testing how effective is the simplest method that banks usually use to identify who their most valuable customers are, which is based on simple variables: customer's Age, Income, Assets and Debt. With an *Explanatory Regression Model* and a *Predictive Regression Model*, we conclude that these variables explain only 51% of the same year's customer's Gross Income, and only 39% of the next year's Gross Income. With these conclusions, banks may find relevant to consider alternative methods for evaluating their customers' value.

Nevertheless, these Regression Models have limitations, which may be improved by future research. First, using *Gross Income* as a measure may be misleading, as it is highly influenced by its financial component and may distort the results. Future research may benefit from using an alternative measure less impacted by the financial component or by using only the non-financial component of Gross Income. Second, the linear regression models did not perfectly conform with the linear model assumptions, so other types of models may be tested in future research. We also suggest for Future research the extension of these regression models by including more variables, as a way of better explaining and predicting Gross Income. Such variables may be customers' product ownership, product usage and lagged variables (to capture the historical evolution of the relationships).

Finally, we presented a *Migration Model* based on *Markov Chains* to estimate CLV, which may be the basis for an alternative model to be used by banks to identify their most valuable customers. The Migration Model was concluded to be very challenging to put in practice in

retail banking, for two main reasons: 1- retail banks offer many products, so there are thousands of possible *engagements levels* 2- pricing is non-standard and very customer-specific.

In order to exemplify the application of this methodology, 968 micro segments (the *Markov States*) were defined, through the intersection of customers' *Age*, *Income* and *Assets+Debt*. The major limitation is the fact that this is a small number of segments and these variables may not be clear value drivers, so the results are not yet much valuable for the bank. In future research, we first recommend defining more segments/states, so that Gross Income within the same segment has a small standard deviation. Also, segments must be based on more customer's attributes, such as their products ownership. Other limitations to be addressed in future research are having a way to validate the accuracy of estimates, and allowing for higher-order Markov Chains (next period's state does not solely depend on the previous state).

Still, the Migration Model application was valuable for the bank in the component of descriptive statistics, namely giving visibility into how customers are distributed among segments, how much Gross Income is generated by each segment, the average Gross Income by segment, and how customers move across segments between two years.

7- References

Berger, Paul D. and Nasr, Nadal I.. 1998. "Customer Lifetime Value: Marketing Models and Applications". Journal of Interactive Marketing, 12(1).

Dwyer, Robert F.. 1997. "Customer Lifetime Valuation to Support Marketing Decision Making". Journal of Direct Marketing, 11(4): 6-13.

Ekinci, Yeliz, Uray, Nimet, Ulengin, Fusun. 2012. "A customer lifetime value model for the banking industry: a guide to marketing actions". European Journal of Marketing, 48(3/4): 761-784

Gupta, Sunil, Lehnamm, Donald R. and Stuart, Jennifer Ames. 2004. "Valuing Customers". Journal of Marketing Research, 41: 7-18.

Fader, Peter S., Hardie, Bruce G. S., Lee, Ka Lok. 2005. ""Counting Your Customers" the Easy Way: An Alternative to the Pareto/NBD Model". Marketing Science, 24(2): 275-284

Haenlein, Michael; Kaplan, Andreas M. and Beeser, Anemone J.. 2007. "A Model to Determine Customer Lifetime Value in a Retail Banking Context". European Management Journal, 25(3): 221-234.

H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

Jackson, Barbara B.. 1985. Winning and Keeping Industrial Customers. Lexington, MA: D.C. Heath and Company.

Jain, Dipak and Singh, Siddhartha S.. 2002. "Customer Lifetime Value Research in Marketing: A Review and Future Directions". Journal of Interactive Marketing, 16(2): 34-46.

Kuhn, Max. 2008. Caret package. Journal of Statistical Software, 28(5)

Matlhouse, Edward C. and Blattberg, Robert C.. 2005. "Can we predict Customer Lifetime Value?". Journal of Interactive Marketing, 19(1)

Pfeifer, Philip E. and Carraway, Robert L.. 2000. "Modeling Customer Relationships as Markov Chains". Journal of Interactive Marketing, 14(2): 43-52.

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Reinartz, W. J. and Kumar, V.. 2000. "On the Profitability of Long Lifetime Customers: An Empirical Investigation and Implications for Marketing".

Rust, Roland T.; Lemon, Katherine N. and Zeithaml, Valarie A.. 2004. "Return on Marketing: Using Customer Equity to Focus Marketing Strategy". Journal Marketing, 68: 109-127.

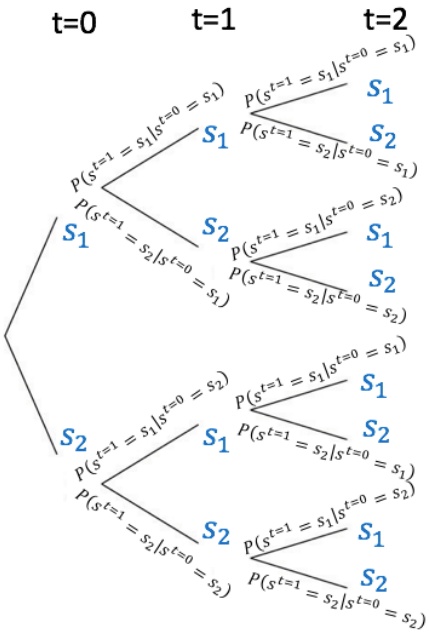
Stahl, Heinz K.; Matzler, Kurt and Hinterhuber, Hans H.. 2003. "Linking customer lifetime value with shareholder value". Industrial Marketing Management, 32: 267-279.

Vafeiadis Thanasis; Diamantaras, Kostas and Chatzisavvas, Konstantinos Ch.. 2015. "A comparison of machine learning techniques for customer churn prediction". Simulation Modelling Practice and Theory, 55: 1-9.

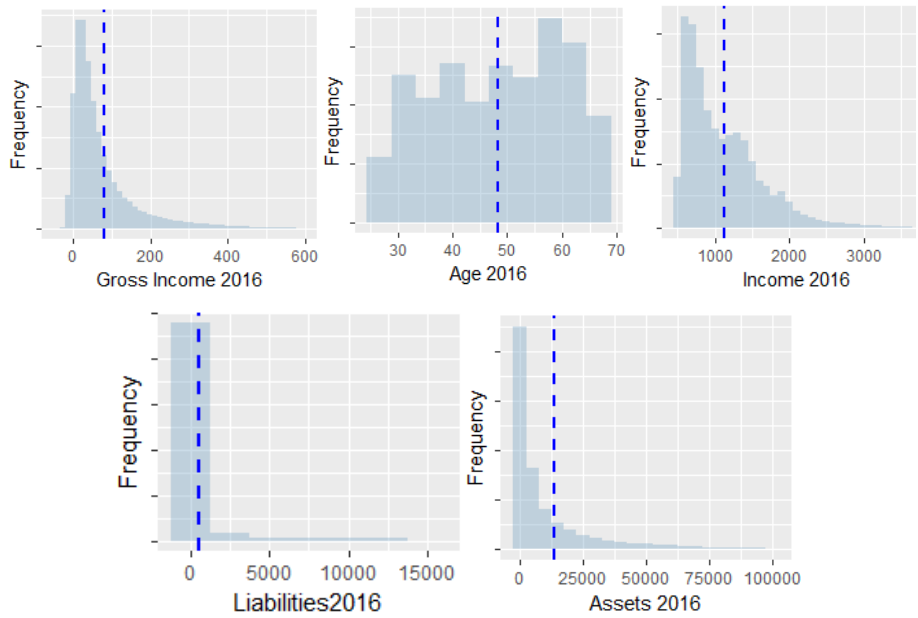
Varian, Hal R.. 2013. "Big Data: New Tricks for Econometrics".

White, D. J.. 1993. "A Survey of Applications of Markov Decision Processes". Journal Operational Research Society, 44(11): 1073-1096.

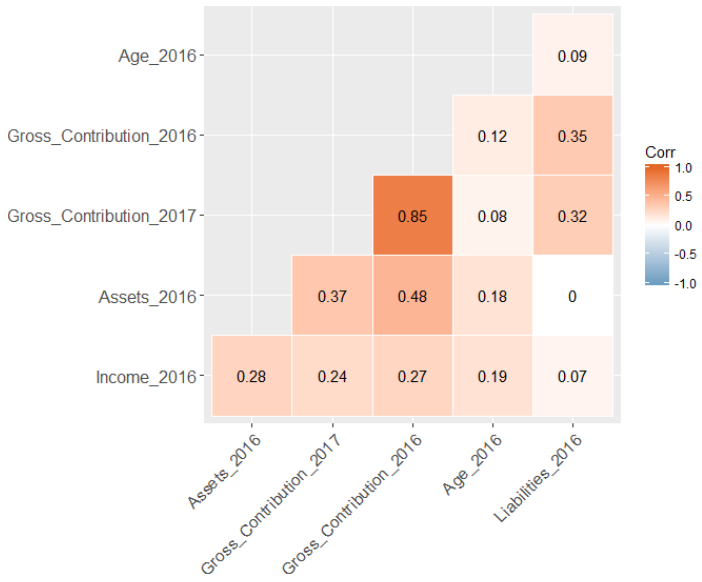
8- Appendices



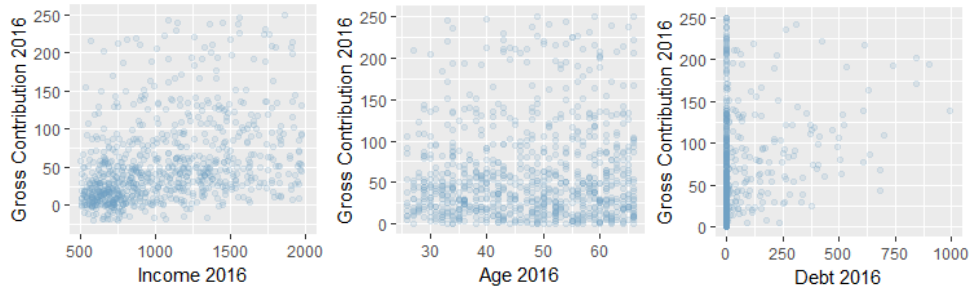
Appendix 1- Example of a tree representing a Markov Chain with two states and 2 periods

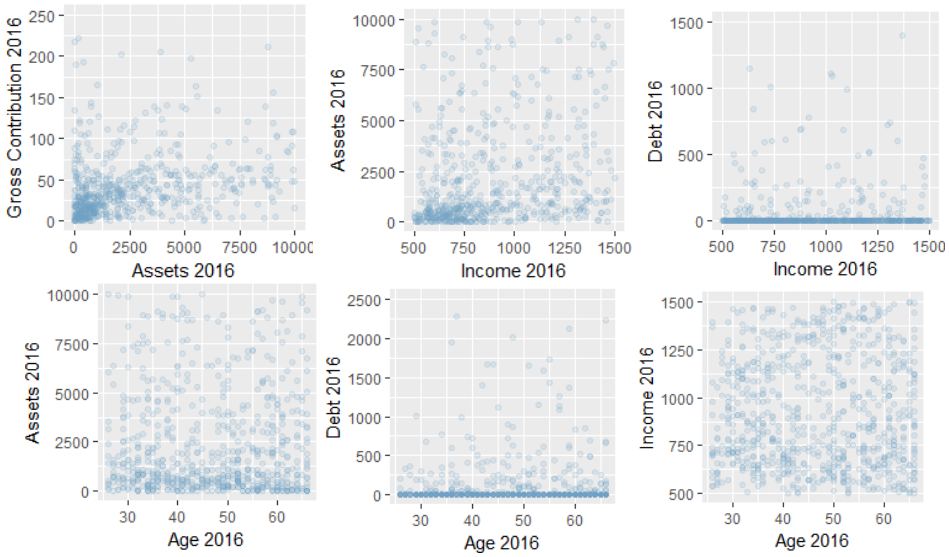


Appendix 2- Histograms presenting the distribution of Gross Income, Age, Income, Debt and Assets for 2016. The mean is represented by the vertical line



Appendix 3- Correlation matrix between the 5 variables of Dataset B



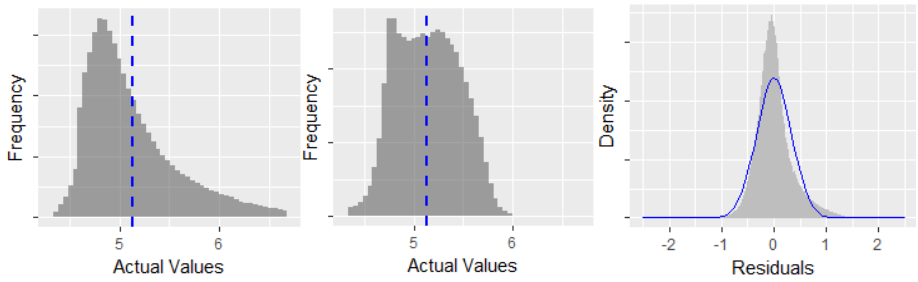


Appendix 4- Scatter plots of the relationships between two varibales, for a selected set of variables (sample fo 1000 observations)

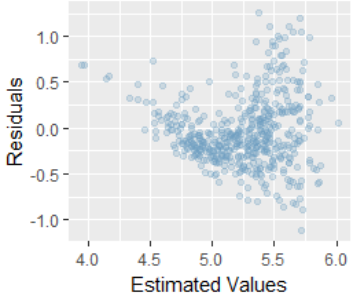
```
Coefficients:
                                 Estimate Std. Error  t value Pr(>|t|)
(Intercept)                     5.687e+00  4.166e-02  136.526  < 2e-16 ***
Age_2016                        2.282e-03  6.808e-04    3.351 0.000804 ***
Assets_2016                    -2.458e-01  3.096e-03  -79.378  < 2e-16 ***
Liabilities_2016                8.872e-02  2.043e-03   43.416  < 2e-16 ***
Income_2016                    -1.451e-01  6.361e-03  -22.809  < 2e-16 ***
Age_2016:Assets_2016            9.875e-04  1.751e-05   56.407  < 2e-16 ***
Age_2016:Liabilities_2016      -2.731e-05  1.143e-05   -2.390 0.016840 *
Age_2016:Income_2016           -1.348e-03  1.037e-04  -12.998  < 2e-16 ***
Assets_2016:Liabilities_2016 -1.384e-02  4.535e-05 -305.178  < 2e-16 ***
Assets_2016:Income_2016         4.238e-02  4.463e-04   94.958  < 2e-16 ***
Liabilities_2016:Income_2016  1.177e-02  3.049e-04   38.606  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3322 on 517898 degrees of freedom
Multiple R-squared:  0.5054,    Adjusted R-squared:  0.5054
F-statistic: 5.292e+04 on 10 and 517898 DF,  p-value: < 2.2e-16
```

Appendix 5- Regression Output for the Explanatory Regression Model. Depedent variable: Gross_Income_2016



Apprendix 6- Histograms presenting the distribution of the actual, estimated values, and residuals

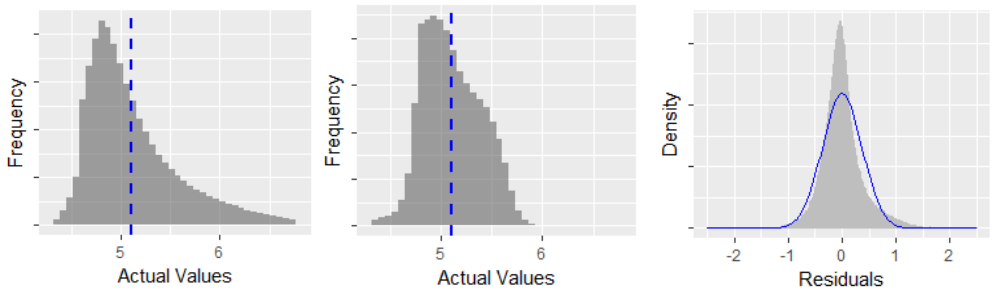


Appendix 7- Scatter plot of the residuals against estimated values and actual values (sample of 500 obs.)

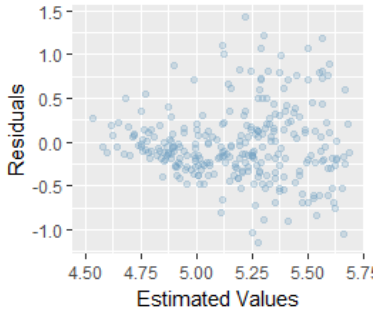
```
Coefficients:
                                 Estimate Std. Error  t value Pr(>|t|)
(Intercept)                     5.385e+00  4.571e-02  117.811  < 2e-16 ***
Age_2016                       -5.597e-03  7.471e-04   -7.492 6.79e-14 ***
Assets_2016                    -1.730e-01  3.397e-03  -50.908  < 2e-16 ***
Liabilities_2016                8.603e-02  2.242e-03   38.367  < 2e-16 ***
Income_2016                    -5.722e-02  6.981e-03   -8.197 2.47e-16 ***
Age_2016:Assets_2016            1.014e-03  1.921e-05   52.772  < 2e-16 ***
Age_2016:Liabilities_2016      -8.642e-05  1.254e-05   -6.893 5.47e-12 ***
Age_2016:Income_2016           -4.485e-04  1.138e-04   -3.940 8.13e-05 ***
Assets_2016:Liabilities_2016 -1.057e-02  4.976e-05 -212.356  < 2e-16 ***
Assets_2016:Income_2016         2.731e-02  4.898e-04   55.757  < 2e-16 ***
Liabilities_2016:Income_2016    8.594e-03  3.346e-04   25.688  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3646 on 517898 degrees of freedom
Multiple R-squared:  0.3874,    Adjusted R-squared:  0.3874
F-statistic: 3.276e+04 on 10 and 517898 DF,  p-value: < 2.2e-16
```

Appendix 8- Regression output for the Explanatory Regression Model.



Appendice 9- Histograms presenting the distribution of the actual, estimated values, and residuals



Appendix 10- Scatter plot of the residuals against estimated values and actual values (sample of 300 obs.)

A+D \ Inc.	1	2	3	4	5	6	7	8	9	10	11	Total
1	13,3%	1,2%	1,0%	0,8%	0,8%	0,6%	0,4%	0,4%	0,6%	0,3%	0,1%	19,5%
2	2,8%	0,3%	0,4%	0,4%	0,3%	0,3%	0,3%	0,3%	0,6%	0,3%	0,1%	6,1%
3	3,4%	0,3%	0,3%	0,3%	0,3%	0,3%	0,3%	0,4%	0,8%	0,4%	0,2%	6,7%
4	4,1%	0,2%	0,2%	0,3%	0,3%	0,2%	0,2%	0,3%	0,6%	0,4%	0,3%	7,2%
5	4,4%	0,2%	0,2%	0,2%	0,3%	0,2%	0,2%	0,3%	0,5%	0,4%	0,3%	7,3%
6	4,6%	0,2%	0,3%	0,3%	0,3%	0,3%	0,3%	0,3%	0,6%	0,5%	0,3%	8,0%
7	4,8%	0,3%	0,3%	0,3%	0,3%	0,3%	0,3%	0,4%	0,7%	0,6%	0,4%	8,7%
8	4,9%	0,3%	0,3%	0,3%	0,4%	0,4%	0,4%	0,5%	0,9%	0,7%	0,6%	9,6%
9	4,2%	0,3%	0,3%	0,3%	0,4%	0,4%	0,4%	0,6%	1,0%	0,8%	0,8%	9,6%
10	3,2%	0,2%	0,2%	0,2%	0,3%	0,4%	0,5%	0,7%	1,1%	1,0%	1,2%	9,0%
11	2,6%	0,1%	0,1%	0,1%	0,2%	0,3%	0,3%	0,5%	1,0%	1,0%	1,9%	8,2%
Total	**52,4%**	**3,6%**	**3,7%**	**3,5%**	**3,9%**	**3,7%**	**3,6%**	**4,7%**	**8,4%**	**6,4%**	**6,1%**	**100%**

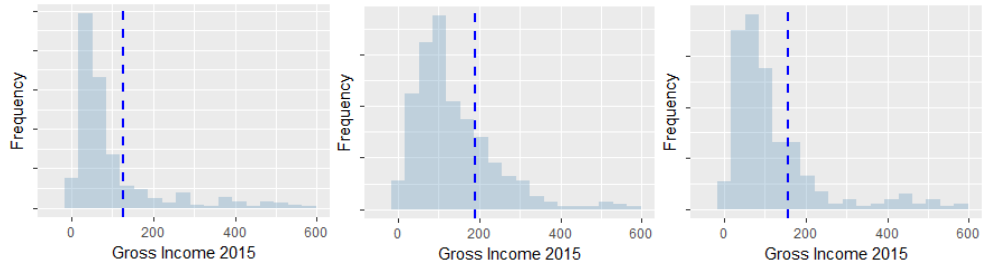
Appendix 11- Percentage of total number of customers in each segment (age group 51-55)

A+D	Inc. 1	2	3	4	5	6	7	8	9	10	11	Total
1	4,2%	0,2%	0,2%	0,2%	0,2%	0,1%	0,1%	0,1%	0,2%	0,1%	0,1%	5,7%
2	1,5%	0,1%	0,1%	0,1%	0,1%	0,1%	0,1%	0,2%	0,3%	0,2%	0,1%	2,7%
3	2,1%	0,1%	0,1%	0,1%	0,1%	0,2%	0,2%	0,2%	0,5%	0,3%	0,2%	4,2%
4	2,8%	0,1%	0,1%	0,1%	0,2%	0,2%	0,2%	0,3%	0,6%	0,5%	0,3%	5,4%
5	2,9%	0,2%	0,2%	0,2%	0,2%	0,2%	0,2%	0,3%	0,6%	0,5%	0,4%	5,9%
6	3,5%	0,2%	0,2%	0,2%	0,3%	0,3%	0,3%	0,4%	0,7%	0,6%	0,5%	7,2%
7	4,7%	0,3%	0,3%	0,3%	0,4%	0,4%	0,4%	0,5%	0,8%	0,7%	0,7%	9,5%
8	5,9%	0,4%	0,4%	0,5%	0,5%	0,6%	0,6%	0,7%	1,3%	1,0%	0,9%	12,8%
9	6,9%	0,4%	0,5%	0,5%	0,6%	0,7%	0,7%	1,0%	1,6%	1,4%	1,4%	15,8%
10	6,6%	0,3%	0,4%	0,4%	0,5%	0,7%	0,7%	1,1%	1,6%	1,7%	2,0%	16,0%
11	7,3%	0,2%	0,3%	0,3%	0,3%	0,4%	0,4%	0,5%	1,5%	1,4%	2,0%	14,7%
Total	**48,5%**	**2,6%**	**2,9%**	**3,0%**	**3,4%**	**3,9%**	**3,9%**	**5,3%**	**9,7%**	**8,3%**	**8,5%**	**100%**

Appendix 12- Percentage of total Gross Income generated by each segment (age group 51-55)

A+D	Inc. 1	2	3	4	5	6	7	8	9	10	11	Total
1	0,32	0,17	0,18	0,20	0,22	0,26	0,29	0,33	0,39	0,44	0,61	0,30
2	0,52	0,27	0,29	0,31	0,33	0,37	0,39	0,51	0,47	0,50	0,49	0,45
3	0,64	0,46	0,47	0,46	0,51	0,59	0,62	0,67	0,66	0,70	0,76	0,62
4	0,68	0,58	0,57	0,56	0,66	0,77	0,84	0,99	0,93	1,06	1,27	0,76
5	0,67	0,64	0,70	0,71	0,81	1,00	0,93	1,05	1,18	1,17	1,51	0,81
6	0,76	0,71	0,92	0,85	0,99	1,07	1,04	1,19	1,08	1,23	1,59	0,90
7	0,97	1,07	1,02	1,11	1,25	1,18	1,30	1,23	1,21	1,30	1,50	1,09
8	1,21	1,40	1,35	1,36	1,31	1,46	1,46	1,41	1,50	1,42	1,63	1,32
9	1,65	1,55	1,69	1,66	1,44	1,64	1,58	1,64	1,60	1,70	1,74	1,64
10	2,03	1,77	1,88	1,61	1,71	1,84	1,61	1,64	1,37	1,69	1,71	1,78
11	2,81	2,14	2,28	2,44	1,60	1,45	1,36	0,98	1,48	1,38	1,06	1,80
Total	**0,93**	**0,70**	**0,80**	**0,85**	**0,89**	**1,04**	**1,09**	**1,13**	**1,15**	**1,29**	**1,39**	**1,00**

Appendix 13- Average Gross Income of each segment (total average is the reference). Age group 51-55



Appendix 14- Histogram of the distribution of Gross Income for three of of the 121 segments of the age group 51-55. The first is for Income Level 5 and Assets+Debt 5, the second for Income 6 and Assets+Debt 6 and the third for Income 7 and Assets+Debt 5

t	t+1 0	1--1	1--2	1--3	1--4	1--5	1--6	1--7	1--8	1--9	1--10	1--11	2--1	2--2	2--3	2--4	2--5	2--6	2--7	2--8	2--9	2--10	2--11	3--1	3--2	3--3	3--4	3--5	3--6	3--7	3--8	3--9	3--10	3--11
1--1	10%	73%	5%	3%	1%	1%	1%	0%	0%	0%	0%	0%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
1--2	4%	29%	36%	14%	5%	2%	1%	1%	0%	0%	0%	0%	1%	1%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%
1--3	4%	15%	13%	43%	13%	4%	2%	1%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
1--4	3%	6%	5%	11%	51%	12%	3%	2%	1%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
1--5	3%	4%	2%	4%	13%	54%	12%	3%	1%	1%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
1--6	2%	2%	1%	1%	3%	13%	59%	11%	2%	1%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
1--7	1%	1%	0%	1%	1%	3%	13%	63%	9%	1%	1%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%
1--8	1%	1%	0%	0%	1%	1%	2%	12%	69%	8%	1%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%
1--9	1%	0%	0%	0%	0%	1%	1%	2%	12%	70%	6%	1%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%
1--10	1%	0%	0%	0%	0%	0%	1%	1%	2%	12%	73%	5%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%
1--11	0%	0%	0%	0%	0%	0%	0%	1%	1%	2%	8%	83%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
2--1	1%	14%	2%	0%	0%	0%	0%	0%	0%	0%	0%	49%	5%	2%	1%	0%	0%	0%	0%	0%	0%	0%	15%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
2--2	1%	7%	3%	2%	1%	1%	0%	0%	0%	0%	0%	19%	21%	9%	2%	1%	1%	0%	1%	0%	0%	0%	7%	11%	5%	1%	1%	0%	0%	0%	0%	0%	0%	0%
2--3	0%	4%	3%	5%	2%	0%	0%	0%	0%	0%	0%	9%	11%	17%	8%	3%	0%	1%	0%	0%	0%	0%	5%	6%	7%	4%	0%	1%	0%	0%	0%	0%	0%	0%
2--4	1%	3%	1%	3%	7%	3%	1%	0%	0%	0%	0%	4%	3%	7%	20%	9%	1%	1%	1%	0%	0%	0%	1%	2%	3%	10%	3%	1%	0%	0%	0%	0%	0%	0%
2--5	0%	1%	0%	1%	4%	7%	2%	0%	0%	0%	0%	1%	0%	1%	7%	28%	11%	1%	1%	0%	0%	0%	1%	0%	1%	2%	9%	3%	1%	0%	0%	0%	0%	0%
2--6	0%	1%	0%	0%	1%	4%	12%	4%	0%	0%	0%	1%	1%	0%	2%	10%	27%	6%	1%	0%	0%	0%	0%	0%	0%	1%	3%	9%	3%	0%	0%	0%	0%	0%
2--7	0%	0%	0%	0%	0%	1%	3%	13%	2%	0%	0%	0%	0%	0%	0%	1%	8%	26%	6%	0%	0%	0%	0%	0%	0%	0%	1%	4%	12%	2%	0%	0%	0%	0%
2--8	0%	0%	0%	0%	0%	0%	1%	3%	15%	1%	0%	0%	0%	0%	0%	1%	1%	7%	35%	4%	0%	0%	0%	0%	0%	1%	0%	0%	2%	13%	2%	0%	0%	0%
2--9	0%	0%	0%	0%	0%	0%	0%	1%	2%	19%	1%	0%	0%	0%	0%	0%	0%	1%	6%	34%	2%	0%	0%	0%	0%	0%	0%	0%	0%	2%	16%	1%	0%	
2--10	0%	0%	0%	0%	0%	0%	0%	0%	0%	3%	16%	0%	0%	0%	0%	0%	0%	0%	0%	1%	6%	40%	2%	0%	0%	0%	0%	0%	0%	0%	2%	15%	2%	
2--11	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	5%	19%	0%	0%	0%	0%	1%	0%	0%	1%	1%	3%	36%	0%	0%	0%	0%	0%	0%	0%	0%	0%	4%	16%
3--1	1%	6%	1%	0%	0%	0%	0%	0%	0%	0%	0%	11%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	49%	6%	1%	1%	0%	1%	0%	0%	0%	0%	0%
3--2	1%	3%	1%	1%	1%	0%	0%	0%	0%	0%	0%	4%	6%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	16%	23%	8%	3%	2%	1%	0%	0%	0%	0%	0%
3--3	0%	2%	2%	2%	1%	0%	0%	0%	0%	0%	0%	2%	3%	5%	2%	0%	0%	0%	0%	0%	0%	0%	0%	6%	10%	20%	9%	2%	1%	0%	0%	0%	0%	0%
3--4	0%	2%	1%	2%	2%	1%	0%	0%	0%	0%	0%	1%	1%	3%	5%	2%	0%	0%	0%	0%	0%	0%	0%	3%	4%	12%	22%	10%	3%	1%	0%	0%	0%	0%
3--5	1%	1%	0%	1%	3%	3%	2%	1%	0%	0%	0%	1%	0%	1%	2%	5%	2%	1%	0%	0%	0%	0%	0%	1%	1%	2%	10%	25%	7%	2%	0%	0%	0%	0%
3--6	0%	0%	0%	0%	0%	2%	5%	1%	0%	0%	0%	0%	0%	0%	0%	1%	2%	7%	2%	0%	0%	0%	0%	1%	1%	1%	2%	11%	34%	8%	0%	0%	0%	0%
3--7	0%	0%	0%	0%	0%	0%	3%	6%	1%	0%	0%	0%	0%	0%	0%	0%	1%	5%	2%	0%	0%	0%	0%	0%	0%	0%	0%	2%	8%	36%	6%	0%	0%	0%
3--8	0%	0%	0%	0%	0%	0%	0%	2%	7%	1%	0%	0%	0%	0%	0%	0%	0%	1%	7%	1%	0%	0%	0%	0%	0%	0%	0%	1%	1%	7%	39%	3%	1%	0%
3--9	0%	0%	0%	0%	0%	0%	0%	0%	2%	9%	1%	0%	0%	0%	0%	0%	0%	0%	0%	1%	7%	0%	0%	0%	0%	0%	0%	0%	0%	1%	7%	38%	3%	0%
3--10	0%	0%	0%	0%	0%	0%	0%	0%	0%	3%	7%	1%	0%	0%	0%	0%	0%	0%	0%	0%	2%	8%	0%	0%	0%	0%	0%	0%	0%	0%	0%	8%	40%	1%
3--11	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	16%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	6%	0%	0%	0%	0%	0%	0%	0%	0%	0%	5%	36%

Appendix 15- Subset of the transition matrix for the age group 51-55. For example, "2--1" represents Income Level 2 and Assets+Debt Level 1

A+D \ Inc.	1	2	3	4	5	6	7	8	9	10	11
1	0,99	0,87	0,88	0,95	1,07	1,25	1,36	1,66	1,76	2,00	2,24
2	1,37	1,12	1,15	1,18	1,31	1,43	1,61	1,86	1,95	2,22	2,64
3	1,61	1,42	1,43	1,46	1,57	1,79	1,97	2,25	2,34	2,60	2,90
4	1,85	1,77	1,75	1,83	1,97	2,16	2,34	2,59	2,73	3,03	3,57
5	2,03	2,12	2,09	2,21	2,38	2,57	2,72	2,95	3,12	3,38	3,97
6	2,35	2,40	2,49	2,62	2,80	2,90	3,08	3,29	3,42	3,70	4,19
7	2,82	3,01	3,09	3,23	3,34	3,41	3,48	3,64	3,73	3,90	4,43
8	3,52	3,74	3,80	3,89	3,90	4,00	4,08	4,11	4,21	4,32	4,66
9	4,53	4,59	4,60	4,56	4,43	4,55	4,54	4,59	4,60	4,75	4,93
10	5,61	5,29	5,19	5,04	4,98	4,96	4,77	4,62	4,49	4,80	4,87
11	7,40	6,35	6,30	5,92	5,36	4,90	4,39	4,03	4,32	4,18	3,75

Appendix 16- 3-year CLV estimation for each customer segment (age group 51-55)