| European Ph.D. thesis |

A MODEL OF CUSTOMER LIFETIME VALUE AND *EX POSTE* VALUE-BASED SEGMENTATION

AN APPLICATION IN A FINANCIAL MULTI-SERVICES RETAILER BASED ON PROBABILISTIC AND DATA MINING MODELS

Doctoral candidate:

Antonia Mª Estrella-Ramón

University of Almería (Spain) - International Campus of Excellence in Agrifood (ceiA3)

Promoter:

Manuel Sánchez-Pérez

University of Almería (Spain) - International Campus of Excellence in Agrifood (ceiA3)

Co-Promoters:

Gilbert Swinnen

Koen VanHoof

University of Hasselt (Belgium)

- Almería, February 2014 -

Agradecimientos

Quisiera expresar mi agradecimiento en primer lugar a la Junta de Andalucía por concederme en 2009 una beca predoctoral para áreas de conocimiento deficitarias en docencia, la cual me ha permitido llevar a cabo este proyecto con una dedicación completa, así como a la Universidad de Almería por acogerme. Agradecer también el apoyo de la empresa suministradora de los datos (entidad que ofrece servicios financieros a nivel nacional), ya que sin ellos este proyecto no se hubiera podido hacer realidad. Seguidamente a mis tres directores: Prof. Dr. D. Manuel Sánchez Pérez, Prof. Dr. D. Gilbert Swinnen y Prof. Dr. D. Koen VanHoof, por brindarme la oportunidad de realizar un trabajo de investigación con proyección internacional. Es por ello que también he de agradecer el apoyo recibido de la Universidad de Hasselt (Bélgica), a la que pertenecen mis dos codirectores y como no, a Dña. Elsa Swinnen, quien me hizo sentir como en casa cada vez que he visitado esta universidad. También agradecer su apoyo desde el primer día que entré en la Facultad de Ciencias Económicas y Empresariales, en general al Departamento de Economía y Empresa, y en particular al resto de profesores del Área de Comercialización e Investigación de Mercados. Confiaron en mis posibilidades como investigadora y docente y gracias a ello he podido desarrollar ambas facetas. En especial quisiera darle las gracias a la profesora Dña. Mª de los Ángeles Iniesta Bonillo, por ser la primera persona que vio en mi posibilidades para dedicarme a la docencia universitaria y la investigación. Asimismo, quisiera agradecerle al Prof. Dr. D. David Jiménez Castillo el haber dedicado tantas horas a enseñarme a investigar, a corregirme lo que escribía y darme tantos buenos consejos, a la Prof. Dr. Dña. Raquel Sánchez Fernández, por su calidad humana que ha conllevado el preguntarme casi cada día cómo estaba y a la Prof. Dr. Dña. Cristina Segovia López, igualmente por su calidad humana y su trato cercano. Agradecer el apoyo del Prof. Dr. D. Antonio Salmerón Cerdán. Sus consejos sobre estadística Bayesiana han sido cruciales para el desarrollo de este trabajo. Finalmente, agradecer su apoyo a la Prof. María del Mar Gálvez Rodríguez y a Dña. Laura Trujillo Dote, con quienes los *papers*, los análisis estadísticos, las revisiones teóricas y las clases han sido mucho más llevaderos.

En el ámbito más personal, agradecer la tarea de mi *personal coach*, Dña. Elena Martínez Bissetti, quien me ayudó a ver las cosas "desde otra perspectiva". Agradecer también la eterna espera de mis amigas, la labor de mis padres y en general de mi familia y la paciencia de Jesús. Gracias a todos ellos por sus consuelos y sus palabras de apoyo, ánimo y fuerza.

A todos ellos, y a muchos más, muchas gracias.

Antonia Mª Estrella Ramón

Almería, January 20, 2014

This dissertation has been supported by the Consejería de Innovación, Ciencia y Empresa of the Junta de Andalucía through the research scholarship: "Formación de personal docente e investigador predoctoral en áreas de conocimiento deficitarias (2009)".

Dedicado a la honradez, al esfuerzo, a la dedicación y a la entrega sin límite y muy en especial a todos aquellos que las ponen en práctica sin tomar atajos en la vida. Confío y deseo que todos ellos pronto consigan sus sueños y nunca, nunca dejen de luchar por todo aquello en lo que creen.

TABLE OF CONTENTS

- Volumen 1 -

|Section| **|Page|**

|Section| **|Page|**

- Volumen 2 -

|Section| **|Page|**

LIST OF FIGURES

- Volumen 1 -

|Title of the figure| **|Page|**

- Volumen 2 -

|Title of the figure| **|Page|**

|Title of the figure|	**|Page|**

|Title of the figure| **|Page|**

|Title of the figure| |Page|

|Title of the figure| **|Page|**

|Title of the figure| **|Page|**

|Title of the figure| **|Page|**

|Title of the figure| **|Page|**

LIST OF TABLES

- Volumen 1 -

|Title of the table| **|Page|**

- Volumen 2 -

|Title of the table| **|Page|**

SUMMARY OF THE RESEARCH

Customer lifetime value (CLV) analysis is a generic term for methodologies that study the net present value of benefits associated with each customer, once he or she has been acquired, after subtracting incremental costs associated with each customer (e.g., marketing, selling, production and service), over his or her entire lifetime with the company. In particular, considering our selected context (a Spanish financial service provider) and the data available (a panel data of customers) described in the following Chapters, we define CLV as the present value of each customer's current and future purchases of banking products. More specifically, CLV is the net present value of the sum of the current and future contribution margins from the customers of the company, which depends on length, depth and breadth of the relationship with each customer, over their lifetimes of operation with the company, taking into account the time value of money using a discount rate to adjust back the predictions about the future to the present.

From a general marketing perspective, this research is motivated by the fact that some recent trends in banking context (e.g., better customer management, focused on strengthening relationships with existing customers through excellence in service quality or develop a customer-centric banking) encourage banks to achieve important challenges in order to stay competitive, especially since the advent of the current financial crisis. Therefore, in order to help to overcome the current economic difficulties we want to offer a tool to manage banking customers. This tool is going to impulse both sides: banks (our CLV model can be considered as a Customer Relationship Management (CRM) tool, which offers an economic assessment of customers) and customers (because this tool can help the bank to understand customer's behaviours and anticipate their needs, which facilitates the relationship and exchange between bank and customers).

From a theoretical point of view, customers and customer relationships have been considered intangible and valuable firm assets since decades. If we link this theoretical proposition with an empirical goal, the result is in line with the analysis of historical records of interactions between the customer and the company in such a way that companies will be able to obtain valuable information that will help them to understand customer's behaviours and anticipate his/her needs, as we have noted previously. This way to proceed ultimately will impact on business performance and in the customer satisfaction with the offer. As firms increasingly see customers as important assets, methods for estimating CLV have been developed as an important strategic marketing tool. For the reasons previously mentioned, we have decided to deal with CLV in a banking context, in order to calculate the value *from* customers as the base for an *ex poste* segmentation scheme. In this way, we present a new research design to obtain a richer customer segmentation taking into account the CLV that each customer brings to the firm (and other sociodemographic information). While traditional customer segmentation is focused on identifying customer groups only using demographics and other attributes (such as attitude and psychological profiles), CLV allows us to undertake customer segmentation in a different way: a value-based segmentation approach.

From a methodological perspective, the objective of this dissertation is to present a new mixture of statistical techniques to model CLV. Customer valuation and customer segmentation problems in marketing have been tackled previously, although in this research we propose a new empirical design that solves both problems in a different and particular way, that is, (1) firstly, estimating CLV_i, where i refers to each customer, and (2) secondly, segmenting customers according to this individual value and other customer characteristics, such as socio-demographic information (i.e., age, gender and income). For these tasks, we have selected certain components and drivers of CLV considered essential in the customer-company relationship. In particular, regarding components of CLV we refer to retention (length dimension), product ownership (breadth dimension), product usage (depth dimension) (for more details about these three dimensions see Bolton *et al*., 2004), contribution margin and also a discount rate to adjust back the predictions about the future to the present. At the same time, we have analysed the underlying behaviours (drivers of CLV) that define length, breadth and depth dimensions that jointly predict CLV (in the following Chapters we give more details about these predictors).

Therefore, using monthly data from a database of 1.357 customers of a Spanish financial services company (a multi-service or multi-product retailer), we present a probability model, in particular a hierarchical Bayesian model, used (1) firstly, to discover those customer characteristics with more potential to predict retention (length dimension), product ownership (breadth dimension) and product usage (depth dimension) and also contribution margin, and (2) secondly, to predict these quantities (using the drivers of CLV) that jointly help us to calculate lifetime value of each customer in our sample (CLV_i, where *i* refers to each customer). Once we identify the most significant predictors of value and we predict length, breadth, depth dimensions and contribution margin, we get the CLV_i (using the formulas enclosed in Chapter

5). As we have mentioned in the previous paragraph, this amount and some socio-demographic information are inputs of an *ex poste* segmentation stage (a value-based segmentation). For this second task, a data mining technique is used, in particular a regression tree. This segmentation is performed in order to identify those groups of customers that are more/less valuable. Our aim pursues to propose the implementation of different strategies to manage different types of customers of the bank. Therefore, both analyses (hierarchical Bayesian models and regression trees) provide excellent opportunities to design a framework that takes into account their interdependencies. Ultimately, these models allow a careful assessment of the contribution of each customer within his/her entire lifetime of operation with the bank and provide a potentially powerful CRM tool to the bank.

Chapter 1. INTRODUCTION

1.1. Introduction

Customers have become the alma mater of any organization, because without them there wouldn't be incomes, benefits and the resulting market value of the company (Gupta and Lehmann, 2003; Gupta and Zeithaml, 2006). The ability to identify profitable customers and build long-term loyal relationships with them is a key factor in the current highly competitive business environment. To achieve this goal, companies have adopted the concept of **Customer Relationship Management (CRM)** as a business strategy in order to retain profitable customers and increase purchases made by them (Jain and Singh, 2002). Under the concept of CRM, customers are not equal and, therefore, it is unreasonable for the company to provide the same offers to all customers. Instead, companies can select only those customers who meet certain profitability criteria based on their individual needs or purchasing behaviours (Dyche and Tech, 2001) to implement **relationship-marketing strategies** with them (Kumar *et al*., 2004). Therefore, a precise evaluation of customer profitability is a crucial element for the success of CRM (Lee and Park, 2005) because this strategy led managers to wonder how they could measure the economic value of a customer in a way that could consider the relationship benefits, the accounting profits that the customer brings to the organization and the prediction of future contributions by the customer. The value of a customer has long been defined with regard to the longevity of his/her historical financial value (Mzoughia and Limam, 2012), but as we are going to show in the following sections, **Customer Lifetime Value measure (CLV)** is a more precise customer value measure that satisfies the previously mentioned considerations (i.e., attempts to account for the anticipated future profitability of each customer relationship).

Continuous advances in information and communication technology play an important role in the previously mentioned customer assessment process, because these advances have allowed companies to collect **large amounts of customer data** at a reduced cost and additionally, they have allowed companies to develop skills to store, share, analyse and transfer valuable information from these data. This trend, coupled with the marketing need to develop key measures to help management control of the business, has caused that such databases are exploited to the maximum (Fader and Hardie, 2009). Currently, these improvements in information technology and the consequently easy availability of transactional data allow companies to perform individual level analysis (customer-by-customer) instead of relying on aggregate survey-based measures (such as satisfaction) (Gupta and Lehman, 2008). Therefore, from these two premises (i.e., customer as the core of every organization and an increasing availability of customer data) it is now more realistic than ever to pass from a transaction-centric approach to a **customer-centric approach in marketing** (Fader *et al*., 2006; Kumar *et al*., 2009; for more details see "*The Path to Customer Centricity*" (Shah *et al*., 2006)). This paradigm is based on the assumption that a satisfied customer becomes a sustainable competitive advantage for the organization, creating a link between these two sides: customer and organization. Analysing the historical records of interactions between the customer and the company, companies will be able to obtain valuable information that will help them to understand customers' behaviours and anticipate their needs, which ultimately will impact on business performance.

This move towards a customer centric approach in marketing has led to an **interest in estimate and understands customers' value or the assessment of customers**. This is an important evolution in various disciplines such as accounting, finance and especially in marketing. Since forties, few companies were beginning to estimate the value of their average customer (The Reporter of Direct Mail Advertising, 1941). Later at the end of the sixties, companies started to use computer technology and the task became more challenging. Companies tried to predict the long-term value of their customers. However, by that time they were the first attempts of this kind of predictive analysis (Petrison *et al*., 1993). For example, Sevin (1965) proposed a method to compute a single customer's profitability by allocating functional groups to each customer and subtracting them from each customer's yearly revenue. This evolution reached marketing discipline in the mid-eighties, which justify the current interest that marketing has been paying to CLV and the related subject of CRM (Haenlein *et al*., 2007). Customer-centric measures are needed in order to achieve the desired improved performance within this customer-centric approach and **CLV is one of the core customer-centric measures** (Verhoef and Lemon, 2013).

Due to the formation of the new concept of CRM in the eighties, a new age of marketing started to grow for which, making a sale was just the beginning of a relationship with a customer, not the end. Relationship marketing constitutes a major shift in marketing theory and practice. Rather than focusing on discrete transactions, it emphasises the establishment, development and maintenance of long-term exchanges (Morgan and Hunt, 1994), because such relationships are thought to be more profitable than short-term relationships as a result of exchange efficiencies between company and customer (Reichheld and Sasser, 1990). An increasing number of companies realised that their most valuable asset was its customer base (Berger *et al*., 2002; Blattberg *et al*., 2001a; Gupta and Lehmann, 2003), which has further increased the focus on managing relationships with them (CRM) and the implementation of **customer relationship approach in marketing** (Reinartz *et al*., 2004). Companies such as Tesco (UK retailer), Capital One (American financial services) and Harrah's Entertainment (American gambling company) are several examples that have successfully used customer data in their marketing strategy to develop and extend customer relationships and enhance customer learning (Verhoef *et al*., 2007).

Such is the importance of the customer as an asset of the company (Srivastava *et al*., 1998, 2001) that even the financial community calls for the inclusion of a set of customer measures in financial reports (Persson and Ryals, 2010). In particular, there is an increasing demand for research to develop more rigorous approaches than existing ones that evidences the relationship between marketing performance and business performance (Gleaves *et al*., 2008) and that justifies the work of the marketing managers to make marketing activities more accountable (Rust *et al*., 2004b). Financial accountability is a key to the success of the firm, because spending without any regard to financial consequences can be disastrous and sound the death knell of the firm (Aravindakshan *et al*., 2004). In this regard, and as we have shown previously, **customer value** measures are critical to assess the performance of business operations, considered as a good approximation of firm value (Gupta *et al*., 2004) and becoming valuable information that should be given to investors (Wiesel *et al*., 2008). In particular, we have identified three stages in the development of customer valuation techniques (Weir, 2008), although for some researchers there is no difference between them (e.g., Mulhern, 1999). The first one pursues only the analysis of the **Customer Profitability (CP)** (e.g., Mulhern, 1999), the second one pursues the analysis of the **Customer Lifetime Value (CLV)** (e.g., Pfeifer and Carraway, 2000) and the third one pursues the analysis of the **Customer Equity (CE)** (e.g., Blattberg and Deighton, 1996).

On one hand, we have compared CLV and CP (for a summary see Table 1), resulting that the first one is the present value of future cash flows, whereas CP refers to an arithmetic calculation of revenues minus costs for a specified period of time (Boyce, 2000; Pfeifer *et al*., 2005). CP is calculated on a single period basis, usually the last economic year (Ryals, 2006) and the time value of money is ignored (Stahl *et al*., 2003). CP is also an accounting summary of events from the present and the past, whereas CLV is forward looking. Then, CLV is a more powerful measure than historic CP analysis (only based on current and past profitability), because CLV also looks at the future potential of the customer (Boyce, 2000; Jain and Singh, 2002). Thus, CP is not forward-looking, not being a good basis for developing marketing strategies (Ryals, 2002). Finally, CLV treats customers as assets and marketing expenditure on them as investment, whereas traditional financial approaches treat marketing as expense, which leads to negative operating margin in the early stages of a high growth company (Gupta, 2009).

Table 1. Comparison between CP and CLV

Customer Profitability	**Customer Lifetime Value**	**References**
CP is an arithmetic calculation of revenues minus costs for a specified period of time.	CLV is the present value of future cash flows.	Boyce (2000); Pfeifer *et al*. (2005)
This measure is calculated on a single period basis, usually the last economic year.	This measure needs several time periods of data to be calculated.	Ryals (2006)
This measure ignores the time value of money.	CLV takes into account a discount rate to transform expected future cash flows into a present value.	Stahl *et al*. (2003)
CP is an accounting summary of events from the present and the past. It is not forward looking.	CLV is forward looking. CLV is a more powerful measure than historic CP analysis. CLV looks at the future potential of the customer.	Boyce (2000); Jain and Singh (2002)
CP is not a good basis for developing marketing strategies.	CLV is a good basis for developing marketing strategies.	Ryals (2002)
CP treats marketing as expense, which leads to negative operating margin in the early stages of a high growth company.	CLV treats customers as assets and marketing expenditure on them as investment.	Gupta (2009)

On the other hand, we have compared CLV and CE (for a summary see Table 2), especially because they are related and sometimes are considered equivalent in the literature. Whereas there is a general agreement on the definition of the first, there are different definitions of CE, in particular (1) some authors define it as the average CLV less acquisition cost (Berger and Nasr, 1998; Blattberg and Deighton, 1996; Blattberg *et al*., 2001a), and (2) other authors propose that the CE of the firm is formed by the CLV's of all the current and potential customers (Zhang *et al*., 2010), which has been found to be a good proxy measure of the equity-market valuation of the firm (Gupta *et al*., 2004). Therefore, compared to CLV, CE is a macro-level measure that can be applied directly to understand equity market reactions to marketing actions (Zhang *et al*., 2010).

Table 2. Comparison between CLV and CE

Customer Lifetime Value	Customer Equity	References
There is a general agreement on the definition of CLV.	There are different definitions of CE:	
	(1) Some authors define it as the average CLV less acquisition cost.	Berger and Nasr (1998); Blattberg and Deighton (1996); Blattberg *et al*. (2001a)
	(2) Other authors propose that the CE of the firm is formed by the CLV's of all the current and potential customers, which has been found to be a good proxy measure of the equity-market valuation of the firm.	Zhang *et al*. (2010); Gupta *et al*. (2004)
It is a micro-level measure.	Is a macro-level measure that can be applied directly to understand equity market reactions to marketing actions.	Zhang *et al*. (2010)

According with the previously shown comparison, firstly we remark that CP is a less powerful measure than CLV and CE, as we can see in the summary of their differences in Table 1. Secondly, CLV and CE are two related measures, therefore when we work with CLV concept, if data are available, it should be reasonable to extend the concept of CLV to CE, especially according to its second definition, in order to get an overall assessment of a firm.

1.2. Importance of CLV estimation for organizations

Nowadays CLV is the most popular customer value measure because:

(i) Many traditional marketing measures (e.g., brand awareness/attitude, market share) are not enough to assess returns of marketing investment, especially in the long-term. This is a serious drawback of these measures because it is meaningful for managers to understand customer value at the individual level to allocate resources accordingly (Zhang *et al*., 2010).

(ii) CLV includes all the elements of customer profitability (Kumar and Shah, 2004; Verhoef and Lemon, 2013).

(iii) CLV is forward-looking (Kumar and Shah, 2004; Verhoef and Lemon, 2013).

(iv) CLV is an essential element of the customer-centric paradigm (Kumar and Shah, 2004; Verhoef and Lemon, 2013).

For these reasons, we focus on the second and third stages of customer valuation (i.e., CLV and CE) to develop this research. CLV is closer to the marketing discipline (research on modelling CLV was one of the MSI research priorities (MSI, 2004)) and it is characterised by more complete analysis, taking into account a greater number of variables (not only financial, as in the first case called Customer Profitability analysis). In particular, as Gupta and Zeithaml (2006) determined, CLV and CE provide a good basis to assess the market value of a firm, moreover marketing decisions based on these observed customer measures improve a financial performance of the firm.

At this point we refer to the question proposed by Gupta and Lehmann (2008): ***"Why do we need CLV in addition to profits, cash flow and other traditional financial metrics?"*** The authors explained that in many businesses CLV, as a marketing productivity measure (Rust *et al*., 2004b), provides greater insight than traditional financial metrics for the following reasons:

(1) The components and drivers of CLV[1] provide important diagnostics about the future health of a business, which may not be obvious from traditional financial metrics.

(2) CLV allows us to assess profitability of individual customers.

(3) It is hard to use traditional financial methods (e.g., discounted cash flow or P/E ratio) to assess the value of high growth companies that currently have a negative cash flow and/or negative earnings. CLV allows us to assess these firms when standard financial methods fail.

(4) CLV provides a structured approach to forecast future cash flows that can be better than using a simple extrapolation approach (e.g., average compound annual growth based on the last five years), as is commonly used in finance.

According to Kumar *et al*. (2009), to achieve the desired profitability, companies could follow two ways or paths: (1) the *conventional path to profitability*, and (2) the *reverse logic framework*.

[1] Persson and Ryals (2010) make a distinction between components and drivers of CE, and by extension, of CLV. First, they point out that the components of CLV and CE are retention rate, cash flows (or alternatively profits) the firm expects to receive from the customer in each future period, and discount rate. Second, they complement CLV concept with its drivers, they are customer perceptions (e.g., satisfaction) and customer behaviours (e.g., purchase frequency).

(1) The *conventional path to profitability* is governed by the launch of innovative products or services as a basis for acquiring new customers. If the company combines these innovations with a rich experience, this leads to satisfaction and satisfaction leads to loyalty, both behavioural loyalty (retention) and attitudinal loyalty (e.g., positive word of mouth). The improved level of retention gives the firm more sales opportunities (by cross-selling and up-selling), increasing profits and profitability, which means that companies have more resources to invest in new innovations, closing the circle of this conventional path to profitability. These companies measure their performance by the number of loyal customers they have, but loyal customers not necessarily are profitable, since the relationship between loyalty and profitability is more complex than it might seem a priori (Kumar *et al*., 2007; Kumar and Rajan, 2009). For this reason there are companies that are beginning to deviate from this conventional path to profitability to (2) the *reverse logic framework* that follows a similar path as the conventional path but in the reverse sense. This reverse path is manifested by the fact that **future customer profitability potential or Customer Lifetime Value (CLV)** forms the basis for any CRM based marketing intervention. Moreover, fundamental to the conceptualization of CRM is the notion of customers' economic value to the business, particularly in a longitudinal sense (Reinartz and Venkatesan, 2008). This notion of long-term economic value gives rise to the measure of CLV.

Therefore, **the estimation of CLV is the key to managing customer relationships (CRM)** (Richards and Jones, 2008), because it is a measure to assess marketing decisions (Blattberg and Deighton, 1996) and to predict customer value of each customer in a company database (Malthouse and Blattberg, 2005; Venkatesan and Kumar, 2004). This is essential to decide about investment in (segments of) customers (Zeithaml *et al*., 2001)) and also to assess the total customer base (Gupta *et al*., 2004) as a summation of CLV predictions of all customers. A large group of researchers have recommended this measure for selecting customers and designing marketing programs (e.g., Reinartz and Kumar, 2003; Venkatesan and Kumar, 2004; Kim *et al*., 2006), because customers selected on the basis of CLV generate more profits than customers selected on the basis of other measures, such as only socio-demographics variables (Reinartz and Kumar, 2003; Venkatesan and Kumar, 2004). More specifically, central to the idea of CRM is the assumption that customers differ in their needs and in the value that they generate for the firm. The way customers are managed should reflect these differences. CRM is therefore not about offering every single customer the best possible service, but about treating customers differently and CLV measure can help us to find such differences.

1.3. Research motivations

Many models have been proposed for measuring CLV since the articles by Dwyer (1989) and Berger and Nasr (1998). All of them have different assumptions under different backgrounds. A deep examination of these studies showed that CLV is often operationalised by considering retention and/or acquisition as the only relevant sources of value (e.g., Gupta *et al*., 2004). We refer here with retention to behavioural loyalty, typically measured in terms of retention rate (e.g., Berger and Nasr, 1998; Fader *et al*., 2005b; Gupta and Lehmann, 2003) and with acquisition to the acquisition of new customers, typically measured as acquisition rate (e.g., Gupta *et al*., 2004; Libai *et al*., 2009; Rust *et al*., 2004a). Many studies have ignored the contribution of other behaviours, such as service/product usage and cross-buying, to business performance (e.g., Blattberg *et al*., 2001a). In particular, about usage (our depth dimension), Jain and Singh (2002) have posited in their future research streams that most CLV models do not include demographics and product usage variables for different product categories. Therefore, it was one of the chosen arguments to include product usage in our CLV model (for more details and arguments see Chapter 4). With respect to cross-buying (our length and breadth dimension), Kamarura *et al*. (2003 p. 47) and later Prinzie and Van den Poel (2008 p. 714), who study the cross-buying of products to discover the hierarchical process of acquisition of financial products, allow us to consider cross-buying as another important component of CLV. In particular, they encourage us to choose cross-buying from the following statement (they did not prove this idea): "*cross-selling is effective for customer retention by increasing switching costs and enhancing customer loyalty, thus directly contributing to customer profitability and lifetime value (CLV)*" (Kamarura *et al*., 2003 p. 47; Prinzie and Van den Poel, 2008 p. 714).

Database marketers are an exception because additionally to retention and acquisition, they have incorporated other sources of value into their calculation of CLV (Hughes, 1996b; Wayland and Cole, 1997). However, many of those studies focus on predicting the future CLV of customers rather than predicting the underlying sources of value or customer purchase behaviours. Inattention to underlying sources of customer value (e.g., not taking into account the level of usage of a service or the additional revenues from customers' cross-buying additional services) can have substantial consequences for the business performance of service companies (Johnson and Selnes, 2004). Additionally, other future research streams are stated and give meaning to our research, such as how can CE strategy be developed from observable behavioural data without requiring customer survey data? In fact, our model does not consider customer's perceptions to estimate CLV, it only considers customer behaviours.

In order to develop this research we have selected certain variables considered essential in the customer-company relationship. In particular, to get a more complete (than existing ones) assessment of customers (in terms of CLV) and to improve CRM strategies, we have selected the following variables as components of customer value: **retention** (length dimension), **product ownership** (breadth dimension), **product usage** or the number of products of each type that each customer owns (depth dimension), **contribution margin** of each customer and the **discount rate** (for more details about these three dimensions see Bolton *et al*., 2004). At the same time, we have analysed the underlying behaviours (in the following Chapters we give more details about these predictor variables) that define these three dimensions and that jointly predict CLV, in general we refer to the following drivers of CLV: recency measures (from the famous triad called RFM), length of the relationship, cross-buying, balance or intensity of products ownership (measured by average monthly assets and average monthly liabilities), adoption of online banking, total quantity of product purchases and one period-lagged variables of product ownership, product usage and contribution margin.

For our purposes, a Spanish financial services retailer gives its support providing a customer's dataset to implement and prove our model. The banking sector, both globally and nationally, is undergoing a major restructuring as a result of the global financial crisis that began in the summer of 2007. In this context and in order to set priorities for the sector, IBM developed during the last quarter of 2009 the "*Spanish Banking Study 2012*" with the cooperation of most Spanish banks (IBM, 2010). Among its findings we highlight the necessity of a better customer management, focused on strengthening relationships with existing customers through excellence in service quality, develop a customer-centric banking, take the situation of concentration and restructuring of the sector to acquire new customers and considering the processes of collecting and managing customer information as a source of competitive advantage. Finally, they also should improve the use of technology to manage the customer experience.

Therefore, we have selected this context because it is a *hot topic* since the advent of the financial crisis. With this research we want to offer a tool to manage banking customers that try to impulse both sides: banks and customers, in order to overcome the current economic difficulties. The retail banking sector becomes predominantly service based and derives higher profits from the creation and retention of long-term relationships with customers, therefore with the analysis of CLV it will be possible to identify the value of each one of such relationships. Another reason that justifies our choice is that among CLV research we have found a common suggestion for further research, i.e., apply the different models in other types of business relationships, especially in the financial services context (Lewis, 2006). Equally important are models that cover a customer's relationships with a portfolio of the products of the company (Rust and Chung, 2006), or in other words, models that deal with different product categories (Jain and Singh, 2002). The purpose of the previously mentioned suggestion is, for example, to examine the effects of marketing dynamics on CLV and CE, i.e., cross-selling between a multi-product brands or products of the firm (Aravindakshan *et al*., 2004). This task constitutes a challenge for this research, because despite the apparent theoretical simplicity of CLV concept, it is fraught with difficulty when applied in practice, in particular in a banking context, where purchase behaviour is rather complex. Customers can purchase more than one service or banking product (there are a large number of (heterogeneous) services/products at their disposal), these purchases are often not independent from each other, it is difficult to assign an amount of profits or contribution margin to each transaction (because of the complex finances in this sector) and additionally, there are different types of transactions and channels available to customers.

Jain and Singh (2002) also call for more research in accurately predicting CLV based on history of usage and prior estimates of CLV, for example using the Bayesian approach (i.e., Gibbs sampling) and providing more accurate estimates of CLV than the traditional regression analysis of historical data. However, some studies have compared the performance of complex versus noncomplex models for customer purchase behaviour and CLV prediction (e.g., Donkers *et al*., 2007; Wübben and Von Wangenheim, 2008; Zhang *et al*., 2010) showing that a model does not necessarily have to be sophisticated in order to precisely forecast a customer value, especially with respect to managerial relevance and applicability. Simple heuristics using initial and repeat purchase data perform well even at the individual level, and sometimes using complex methods instead of simple models does not substantially improve predictive accuracy. This fact is meaningful for marketing researchers who seek research simplicity or attempt to avoid computation intensity. In our case, where we have to develop a model that covers an important number of products and predictors of CLV in an extremely complex context, we had to look for simplicity inside the inherent complexity of the problem. Therefore, Bayesian statistics was the key to solve our problem (Ntzoufras, 2009), in particular using Hierarchical Bayesian models.

1.4. Research goals

A wide range of CLV models with different purposes have been developed until now. For this reason, some authors have collected such models in different classifications, taking into account very specific criteria (for more details see Chapter 3), but here we offer a more general classification according to the main purpose to calculate CLV. After we present this brief classification, we have explained what our own purposes to estimate CLV are.

Research on CLV can be divided into three general streams: (a) modelling and prediction, (b) optimise customer strategies in such a way that CLV is maximised, and (c) research that links CLV and firm performance (Verhoef *et al*., 2007). For each of the previously mentioned categories, different subcategories have been developed to solve particular problems, as it is shown below in Table 3.

Regarding our research goals, the **overall goal** of this research aims to develop an integrated framework for assessing customers in the long-term, based on the Customer Lifetime Value (CLV) and Customer Equity (CE) approaches. In particular, in our research the value that each customer provides to the company will not be limited to benefits of each transaction. It will also cover the future benefits that each customer will provide along his/her relationship with the company (e.g., Kumar and George, 2007), i.e., we develop a predictive approach. From this overall goal, a number of research questions related to **specific goals** are listed in Table 4. Our goals are related to the general sub-categories (in Table 3) a1, a2 and c2.

Table 3. Overview of CLV studies

(a) Modelling and prediction of CLV		
Sub-category	Description	Examples
(a1) *Retention and acquisition*	Relationship marketing, an important theoretical foundation of CLV, emphasises the need for maintaining long-term customer relationships. This fact justifies the primary role of customer retention in the majority of CLV models. Individual customer lifetime profits are traditionally modelled as a function of a customer's lifetime duration, revenue flows over the course of a customer's lifetime and firm costs associated with the marketing exchange, emphasizing the role of customer retention as a driver of CLV.	Malthouse and Mulhern (2008); Reinartz and Kumar (2000); Singh *et al*. (2009)
	Additionally, acquisition of customers is another important component of CLV. Some authors have presented models that capture the complex dynamics of customer acquisition and retention.	Blattberg and Deighton (1996); Gupta *et al*. (2004); Libai *et al*. (2009)
(a2) *Segmentation*	CLV measures are used to segment customers, for example analysing customer value, which comprises current value (past profit contribution), potential value (opportunity to cross-sell) and loyalty (defection probability) and segmenting customers based on these measures.	Hwang *et al*. (2004); Kim *et al*. (2006)
(a3) *Product recommendation*	For example, using a weighted RFM-based method based on CLV and combining AHP (analytic hierarchy process) to evaluate the importance or weight of each RFM variable, clustering (k-means) to group customer by weighted RFM variable and association rule mining techniques to provide product recommendations to each customer group.	Liu and Shih (2005a), (2005b)
(b) Optimise customer strategies in such a way that CLV is maximised		
Sub-category	Description	Examples
(b1) *Marketing resource allocation*	By appropriately allocating marketing resources to targeted customer, firms are better positioned to increase profits. Using CLV as a measure for customer selection and marketing resource allocation enables managers to maintain or improve customer relationships proactively through marketing contacts across various channels and to maximise CLV simultaneously.	Berger and Nasr-Bechwati (2001); Ching *et al*. (2004); Venkatesan and Kumar (2004)
(c) Research that links CLV and firm performance		
Sub-category	Description	Examples
(c1) *Mergers and acquisitions*	Mergers and acquisitions imply the acquisition of the customer base of one company by another one. CLV estimation could provide a link between customer and firm value. It could be useful for a variety of managerial decisions, such as mergers and acquisitions.	Gupta and Lehmann (2003); Gupta *et al*. (2004)
(c2) *CE*	For most firms, CE is bound to be the most important component of the value of the firm. Therefore, understanding how to drive CE is central to the decision making of any firm and formulating a procedure to achieve this CE measure can give the firm an important competitive advantage.	Kumar and Shah (2009); Rust *et al*. (2004a); Wiesel *et al*. (2008)

Table 4. Research questions and specific goals of the current research

Research question	**Specific goal**
Related to the predictors used to predict length (retention) and breadth (cross-buying) dimensions: Are these variables good predictors of these two dimensions of CLV?	Develop an integrated framework for assessing the long-term value of customers (prediction in the long term, e.g., Hui-min *et al.*, 2006), based on **CLV and CE theoretical approaches**, taking into account **customer retention** (length dimension), **product ownership** (breadth dimension), **product usage** (depth dimension) (Bolton *et al.*, 2004), **contribution margin** and **discount rate**. At the same time, we have analysed the underlying behaviours (in the following Chapters we give more details about these drivers or predictor variables) that define these three dimensions or components of CLV and that jointly predict CLV.
Related to the predictors used to predict depth (usage) dimension: Are these variables good predictors of this dimension of CLV?	
Related to the predictors used to predict contribution margin: Are these variables good predictors of future customers' contribution margin?	
Can we rank and order the customers of the bank according to their value?	**Segment the customers of the bank according to their value** (Kim *et al.*, 2006) and other information (socio-demographic): value-based customer segmentation.
What is the overall value of the customer base?	Obtain an overall assessment of the customers (**Customer Equity**), considered the main asset of the company (Gupta *et al.*, 2004).
How could the bank improve CRM?	Identify those customers who are more/less valuable to implement **strategies to manage them according to the value they offer to the company** (Bruhn *et al.*, 2006) using the proposed segmentation scheme**:** *customer retention* (Blattberg *et al.*, 2009) versus *customer divestment* (Mittal and Sarkees, 2006; Reinartz and Kumar, 2002, 2003).

1.5. Design decisions

In this section we explain how we have designed our new approach to assess and segment customers in accordance with the research questions previously shown, highlighting what are our main contributions, both theoretical and empirical.

1.5.1. Theoretical decisions

Firstly, about the **theoretical foundation of this research** we highlight the relatively 'new' role of customers as firm assets (Srivastava *et al.*, 1998; Srivastava *et al.*, 2001). This fact increases the use of methods for estimating CLV as an important strategic marketing tool, which helps firms to improve customer relationship management (CRM). The CUSAMS framework (Bolton *et al.*, 2004) gives support to our theoretical propositions about the importance of customers for the survival of firms and helps us to deal with the measurement and estimation of CLV. It provides a rich perspective to solve our problem (i.e., enables service organizations to make a comprehensive assessment of the value of their customer assets), because it characterises relationships with customers according to three dimensions: (a) length, (b) depth, and (c) breadth of relationships. Thereby, these three dimensions influence CLV (for more details see Chapter 2) and also contribution margin and the discount rate, all of them are called **components of CLV** (Persson and Ryals, 2010). At the same time, we have analysed the underlying behaviours that define these components and that jointly allow to predict CLV (in the following Chapters, especially in Chapter 4, we give more details about these predictors); particularly they are called **drivers of CLV** (Persson and Ryals, 2010). According to our knowledge, these predictors have not been studied together in other previous CLV models. In general, we refer to the following drivers of CLV: recency measures (from the famous triad called RFM), length of the relationship, cross-buying, balance or intensity of products ownership (measured by average monthly assets and average monthly liabilities), adoption of online banking, total quantity of product purchases and one period-lagged variables of product ownership, product usage and contribution margin. Therefore, we propose a global and complete view about customer value.

Previous research has noted that ***customer retention*** (length dimension) and ***product ownership*** (breadth dimension), can contribute in a significant proportion to CLV-CE (Gupta and Zeithaml, 2006). Additionally, ***product usage*** and ***contribution margin*** are also considered important components of CLV (Verhoef and Donkers, 2001; Verhoef, 2004). In particular:

(a) Related to the first component of CLV (length dimension), for the objectives of this research we have selected the following drivers of CLV: ***RFM variables*** and ***length of the relationship*** to measure in a complete way behavioural loyalty (Chang and Tsay, 2004; Kumar *et al*., 2006a; Li *et al*., 2011). On one hand, ***RFM*** is a popular set of three variables: recency (R) or time since the last transaction (or in other words, this variable measures the time difference between the last purchase and the time of analysis), frequency (F) or number of transactions during a time period of calculation and monetary value (M) of transactions. On the other hand, we consider ***length of the relationship*** as another way to measure customer loyalty. This concept refers to the duration of the relationship and customer retention (Bolton *et al*., 2004). Relationship length indicates a level of customer inertia that would be associated with greater loyalty (Bolton *et al*., 2004; Colgate and Lang, 2001; Reinartz and Kumar, 2000). Therefore, we consider

retention (length dimension) as one of the components of CLV and ***behavioural loyalty*** measures as drivers of CLV.

(b) According to the second component of CLV (breadth dimension), we have selected ***cross-selling*** as driver of CLV. Previous research has noted a synergic effect related to this variable, which implies that if the customer-company relationship involves more than one product, as is common in the selected context for this research (i.e., financial services retailer), the ability of the company to cross-sell is configured as a fundamental element for CLV. The ability to cross-sell products produces an effect on the total that is greater than the sum of its parts (Rust and Chung, 2006). This effect has not been extensively tested and documented in the marketing literature and it is configured as a prolific future research stream (Rust and Chung, 2006; Villanueva and Hanssens, 2007). Therefore, we consider ***product ownership*** (breadth dimension) as the second component of CLV and a ***cross-selling*** measure as one of the drivers of CLV.

(c) ***Product usage*** (or the number of banking products of each type that each customer purchases and owns, i.e., depth dimension) is considered the third component of CLV. The dynamic nature of customer relationship is especially important in service firms, such as financial services retailers, because customers' service usage levels have a substantial impact on the long-term profitability of the organization (Bolton and Lemon, 1999), and moreover in CLV (Verhoef, 2004). While a number of researchers have explored the problem of modelling churn in a contractual setting, there is a limited amount of research on modelling the usage under contract (Ascarza and Hardie, 2012). Therefore, ***product usage*** is the third component of CLV (i.e., depth of the relationship).

(d) Finally, ***contribution margin*** and ***discount rate*** are the fourth and the last components of CLV, respectively. For sakes of simplicity researchers usually assume that the margins of products remain constant over time (Verhoef and Donkers, 2001; Verhoef, 2004). However, this assumption is questionable. Following the suggestions of recent authors (Kumar and Shah, 2009; Kumar *et al.*, 2006a; Venkatesan and Kumar, 2004), we also model the contribution margin and we take into account the time value of money using a discount rate to adjust back the predictions about the future contribution margin to the present. Therefore, ***contribution margin*** and ***discount rate*** are the fourth and the last components of our CLV model.

1.5.2. Methodological decisions

Secondly, from a **methodological perspective**, the objective of this dissertation is to present a new mixture of statistical techniques to model CLV. Customer valuation and customer segmentation problems in marketing have been tackled previously, although in this research we propose a new empirical design that solves both problems in a different and particular way, that is, (1) firstly, estimating CLV_i, where i refers to each customer and (2) secondly, segmenting customers according to this individual value and socio-demographic information (age, gender and income). For these tasks, as we have previously mentioned, we have selected certain components and drivers of CLV considered essential in the customer-company relationship (for more details see Figure 1 and Chapter 4).

1.5.2.1. The first methodological stage

We have implemented a two-stage model where the first stage implies the development of a stochastic and behavioural model to estimate and predict individual CLV (CLV_i, where i refers to each customer), based on several drivers and components. Panel data methodologies, such as hierarchical Bayesian model using MCMC, are used: (1) firstly, to discover those drivers with more potential to predict the components of CLV (i.e., retention (length dimension), product ownership (breadth dimension) and product usage (depth dimension) and also contribution margin), and (2) secondly, to predict these components that jointly help us to calculate lifetime value of each customer in our sample (CLV_i). In other words, a hierarchical Bayesian model is developed that jointly predicts a customer's product ownership pattern, product usage pattern and spending pattern (in terms of contribution margin) at each future purchase occasion for a total of 18 products using a sample of 1.357 customers.

Some authors have given detailed overviews and comparisons of the wide range of different approaches that have been used for CLV modelling (e.g., Donkers *et al*., 2007; Gupta *et al*., 2006; Kumar and George, 2007; Ngai *et al*., 2009). In particular, Donkers *et al*. (2007) explained that ***regression type models*** are often used in this context, e.g., linear regression model (Malthouse and Blattberg, 2005; Malthouse and Mulhern, 2008); Probit model (Verhoef and Donkers, 2001); multivariate Probit model (Donkers *et al*., 2007); multivariate Logit model (Prinzie and Van den Poel, 2007). This type of models has the disadvantage that they are smoothing techniques that attempt to describe well the relationship between the predictors and the response but tend to treat heterogeneity as noise (Colombo and Jiang, 1999). Moreover, CLV has been analysed in a substantial number of different domains, ***varying from econometric models to computer science techniques*** (Gupta *et al*., 2006). Some studies have compared the performance of complex versus noncomplex models for customer purchase behaviour and CLV prediction (e.g., Donkers *et al*., 2007; Wübben and Von Wangenheim, 2008; Zhang *et al*., 2010) showing that a model does not necessarily have to be sophisticated in order to precisely forecast a customer value, especially with respect to managerial relevance and applicability. Therefore, this fact is meaningful for marketing researchers who seek research simplicity or attempt to avoid computation intensity, such is our case.

Additionally, it is noteworthy that research on CLV measurement has so far focused on specific contexts, because the data available to a researcher or firm in different contexts might be different. The two types of context generally considered are: non-contractual and contractual (e.g., Reinartz and Kumar 2000, 2003; for more details see Chapter 3, section 3.1.1). Different models for measuring CLV arrive differently at estimates of the expectations of future customer purchase behaviour. In our case, i.e., a financial service retailer, a contractual setting has been chosen as a way to solve our problem (for a deeper understanding of this choice see section 3.1.1).

Getting a solution for our problem implies to develop a model that covers an important number of different and heterogeneous products, because of the inherent characteristic of a financial services retailer to be a multi-service or multi-product provider. Indeed, a financial services retailer is an extremely complex context to apply a model of this type. Additionally, our model has to discover the best drivers of CLV among a wide range of variables. Thus, we had to look for simplicity inside the complexity and Bayesian statistic (Rossi and Allenby, 2003) was the key to solve our problem (instead of using the classical statistical theory). The main difference between the classical statistical theory and the Bayesian approach is that the latter considers parameters as random variables that are characterised by a prior distribution. This prior distribution is combined with the traditional likelihood to obtain the posterior distribution of the parameter of interest on which statistical inference is based. To get this posterior distribution Markov chain Monte Carlo algorithm (MCMC) is used. MCMC allows setting up and estimating complicated models that could not be solved with traditional methods (Ntzoufras, 2009 p.1-2).

Therefore, a ***hierarchical Bayesian model***, with the help of MCMC theory, is used for the implementation of our large model and parameter space using WinBUGS version 1.4.3, available via the WinBUGS project webpage. We have also taken advantage of the ideas found in Borle *et al*. (2008) and Abe (2009b). They call for the inclusion of a rich set of covariates in their

Hierarchical Bayes framework to estimate CLV. Our variables could enrich the estimation of CLV through Hierarchical Bayes approach and at the same time, we could prove if they have the potential to predict the defined components of the value of each customer.

1.5.2.2. The second methodological stage

The second stage implies an *ex poste* segmentation of customers, taking into account CLV model output (CLV_i, where i refers to each customer) and individual socio-demographic information (such as age_i, $gender_i$, $income_{it}$). Some authors propose different approaches firstly, to segment customers and secondly, to get a CLV measure for each group, they are aggregate views of CLV (e.g., Haenlein *et al*., 2007). These models only predict the average CLV at an aggregate level for the entire customer base or for a reduced number of segments of customers, without taking into consideration individual characteristics of customers (i.e., they get an overall CLV or aggregated CLV's for different segments, not individual CLV's). This is a serious drawback, since profitability is usually not distributed uniformly among customers and a primary objective of the lifetime value approach is to identify highly profitable customers in order to keep existing ones (the most profitable ones), and also to identify non-profitable customers in order to avoid investment in them (Tirenni *et al*., 2007). This way to proceed is followed by, for example Haenlein *et al*. (2007), who firstly use decision trees to segment the customer base based on certain drivers of profitability, including customer retention (measured as a dummy variable where 0 means customer inactive and 1 customer activity with the company) and cross-selling (measured by two variables: type and intensity of product ownership; the first one is measured through 11 dummies, each dummy related with each product, where 0 means no ownership of the product and 1 the opposite, and the second one is measured by several variables that collect balances of each customer for each product). After the process of customer segmentation, these authors calculate an average CLV for each of the segments obtained. Maybe they could have obtained more accurate CLV results if they had firstly estimated CLV for individual customers and after that, they had segmented customer base based on the CLV output. More recently, Chan (2008) proceeds differently and identifies customer behaviour using RFM variables from a Nissan automobile retailer to segment customer base through genetic algorithm, and then uses CLV model to assess the proposed segmentation.

Closer to our research proposal, other authors point out that individual CLV estimations can be used as an intermediate step for classification purposes (Bruhn *et al*., 2006; Keiningham *et al*., 2006; Kumar *et al*., 2009), in other words, CLV measure is an interesting input to perform a customer segmentation (Lemon and Mark, 2006). In particular, while traditional segmentation focused on identifying customer groups based on demographics and attributes such as attitude and psychological profiles, CLV undertakes a value-based approach that looks at groups of customers in terms of the revenue they generate and the costs of establishing and maintaining relationships with them (Kumar, 2008b p. 43). Therefore, based on the distribution of CLV's, some authors divide, select and prioritise customers into different profitability segments, for example high CLV, medium CLV and low CLV. In particular, we have found in the literature several segmentation strategies based on CLV. They can be classified into three categories (Kim *et al*., 2006):

(i) Segmentation by using only CLV values, e.g., the customer pyramid (Zeithaml *et al*., 2001);

(ii) Segmentation by using only CLV components, e.g., current value, potential value, loyalty, etc. (Hwang *et al*., 2004); and

(iii) Segmentation by considering both CLV values and other information, e.g., socio-demographic information, transaction history, etc. (Kim *et al*., 2006).

According to our knowledge, all the CLV empirical applications that we have found that propose a clear customer segmentation scheme have in common a major drawback: CLV is not calculated through stochastic and disaggregated models that capture heterogeneity between customers as a first stage of modelling. In this research we want to overcome this drawback through our first empirical stage, getting an accurate CLV measure for each customer. Once we have estimated this individual CLV, we use it to segment the customer base. For this second stage we apply data mining techniques (regression trees) using the software R version 3.0.2, available via the R project for statistical computing webpage.

1.5.3. Research context decisions

The choice of a **financial service retailer** as our research context is motivated by the fact that some recent trends in banking context (e.g., better customer management, focused on strengthening relationships with existing customers through excellence in service quality or develop a customer-centric banking) encourage banks to achieve important challenges in order to stay competitive, especially since the advent of the current financial crisis. Therefore, in order to help to overcome the current economic difficulties we want to offer a tool to manage banking customers. This tool is going to impulse both sides: banks (our CLV model can be considered as a CRM tool which offers an economic assessment of customers) and customers (because this tool can help the bank to understand customer's behaviours and anticipate their needs, which facilitates the relationship and exchange between bank and customers).

Thinking about the retail-banking environment, a model to determine CLV should satisfy at least three conditions (Haenlein *et al*., 2007):

(1) It needs to be able to handle discrete one-off transactions, which occur either only once in a lifetime or in very long purchasing cycles (e.g., mortgages), and continuous revenue streams (e.g., regular account maintenance charges) equally well. This is due to the fact that retail banks generate revenue in two main ways, by gaining a margin on lending and investment activities and by receiving transaction fees for transactions, credit cards, etc. (Garland, 2002).

(2) In order to be easily implementable, it should focus on the assessment of homogeneous segments of customers instead of individual customers (Libai *et al*., 2002). This requires a trade-off between reflecting individual customer characteristics (such as product usage or lifetime phase) and the large size of an average customer base of a bank, where individual assessment would result in a disproportionate effort and an unmanageable complexity. Despite this second condition, following recent authors (Bruhn *et al*., 2006; Keiningham *et al*., 2006; Kumar *et al*., 2009) it would be interesting to estimate individual CLV to be used as an input to perform customer segmentation. This is currently possible due to the enormous advances in information technology that makes easier than some years ago managing large amounts of customer data to perform individual estimations, as we have proposed.

(3) It needs to be easy to understand and parsimonious in nature to ensure its applicability in many business contexts. This specifically implies limiting data requirements to the information available in an average information system of a bank (Haenlein *et al*., 2007).

1.5.4. Concluding remarks

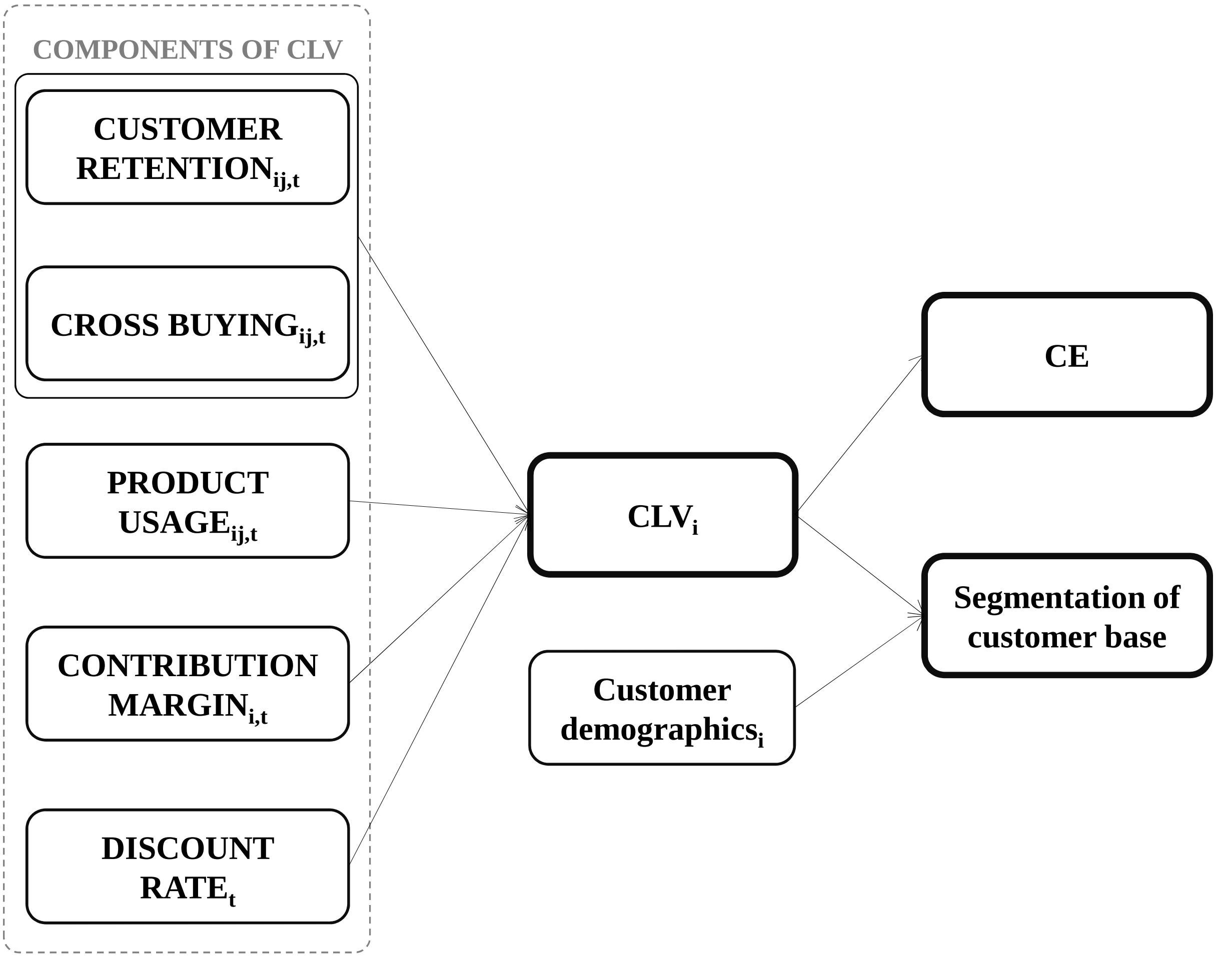
In conclusion, we present a model for the assessment of customers, which we developed in cooperation with a leading Spanish retail bank and which takes the three previously mentioned requirements into account. This model is based on a combination of a Hierarchical Bayesian model to estimate CLV (e.g., Abe, 2009b; Borle *et al*, 2008) and regression tree analysis to answer several important questions:

- **Which drivers of CLV have more potential to predict components of CLV?**
- **What is the long-term value of each customer ($\mathbf{CLV_i}$)?**
- **What is the value of the customer base (CE)?**
- **Which groups of customers are more (less) valuable?**

An in-depth understanding of these questions is of interest and importance to both managers and researchers, since the results can be used as input to marketing decisions, which for example, can contribute to acquire economic returns from customers as an important asset of the company. To serve these interests, a model in Figure 1 is built to develop an integrated framework to estimate individuals CLV's as a basis for segmentation (CLV_i). The model has been estimated and validated using a sample of 1.357 customers with 32.568 datasets (1.357 customers times 24 months of data of each customer) from the collaborative leading Spanish retail bank.

In next sections we review the most relevant theoretical influences of CLV and CE, such as Customer Relationship Management (CRM). We analyse different CLV definitions and propose our own definition taking into account our context and data available. We also proceed in the same way with CE concept. Additionally, we develop an in-depth classification of the most important and referred studies about CLV and CE and we also pay attention to CLV-CE models applied in a similar (e.g., insurance companies) or exactly the same context (i.e., bank), allocating a particular section about CLV and CE in the banking context. Then, we justify the interest of including the variables presented previously in a CLV model. In the methodological section, we discuss the selected context, data available, measures of the variables, statistical models applied and our results. Finally, we discuss whether we have reached the research goals, several management implications and identify limitations and areas for future research.

Figure 1. Summary of the proposed model to assess financial service customers (*)



(*) Where *i* is the customer index, *t* is the time period index, and *j* banking product index.

Chapter 2. THEORETICAL FRAMEWORK

2.1. Theoretical influences of CLV and CE

2.1.1. Customer relationships as valuable assets

From the perspective of the **Resource-Based View (RBV)**, resources that are valuable, rare, inimitable and non-substitutable (Barney, 1991) make it possible for businesses to develop and maintain competitive advantages. Firms need to utilise these resources and competitive advantages to get their superior performance (Collis and Montgomery, 1995; Grant, 1991; Wernerfelt, 1984). Leveraging resources to create and sustain perceived value for the stakeholders of the organization and, in particular customers, has such importance because of the considerable goodness of fit between marketing theorists (Hunt, 2000) and the assumptions of RBV. Many marketing theorists have accepted the RBV approach because it offers a sophisticated explanation of the role that customers play in the creation of value for the firm. In particular, to get this fusion between marketing and RBV theory companies create value for customers identifying resources that are both marketing specific (i.e., they are generated and leveraged in large part through marketing activities) and potentially manifest at least some of the desired RBV attributes (i.e., they are rare, inimitable and non-substitutable). **Market-based assets**, (for a complete definition see Srivastava *et al*., 1998) meet both criteria, allowing customers and their relationships with the firm to be treated as critical resources that contribute to competitive advantage for the firm and which should be developed, augmented, leveraged and valued in a similar way to the traditional resources of the firm.

Furthermore, customers and customer relationships have been considered intangible and valuable firm assets since decades (e.g., Anderson *et al*., 1994; Bursk, 1966; Cravens *et al*., 1997; Gupta *et al*., 2006; Levitt, 1983). In particular, for Srivastava *et al*. (2001) customers are considered **relational market-based assets** of companies. Srivastava and his collegues distinguish between two groups of market-based assets that are essential for firms to get this superior performance: (i) *relational market-based assets* and (ii) *intellectual market-based assets*.

(i) *Relational market-based assets*, including here relationship with stakeholders, e.g., customers, channels, strategic partners, providers of complementary goods and services, outsourcing agreements, networks and eco-system relationships.

(ii) *Intellectual market-based assets*, or types of knowledge a firm possesses about its competitive environment.

Among the first group of these market-based assets (i.e., ***relational market-based assets***), Srivastava *et al*. (2001) highlight the importance of **relationships with customers**. The important role of relationships makes it possible that **Customer Relationship Management (CRM)** becomes a major shift in marketing theory and practice. Rather than focusing on discrete transactions, it emphasises the establishment, development and maintenance of long-term exchanges (Morgan and Hunt, 1994), because such relationships are thought to be more profitable than short-term relationships as a result of exchange efficiencies between company and customer (Reichheld and Sasser, 1990). This paradigm is based on the assumption that a satisfied customer becomes a sustainable competitive advantage for the organization, creating a link between these two sides: customer and organization. Therefore, analysing the historical records of interactions between the customer and the company, companies will be able to obtain valuable information that will help them to understand behaviours of customers and anticipate their needs, which ultimately will impact on business performance.

2.1.2. Customer Relationship Management in the customer valuation framework

The concepts of **market orientation**[2] and **relationship marketing**[3] converge in a business strategy based on **Customer Relationship Management (CRM)**. CRM is defined as the management of mutually beneficial relationships from the perspective of the seller (LaPlaca, 2004 p.463), which benefits all those in the relationship (Mitussis *et al*., 2006). CRM posits that during cooperative and collaborative relationships, value is created for the customer and the firm. In other words, CRM is the enterprise approach aimed at understanding and influencing customer behaviour in order to improve customer acquisition, customer retention, customer loyalty and customer profitability (Swift, 2001). If a more specific definition of CRM is required, Payne and Frow (2005 p. 168), specify it as: "*a strategic approach that is concerned with creating improved shareholder value through the development of appropriate relationships with key customers and*

[2] A *market orientation* is that culture that (1) places the highest priority on the profitable creation and maintenance of superior customer value while considering the interests of other key stakeholders; and (2) provides norms for behaviour regarding the organizational development of and responsiveness to market information (Slater and Narver, 1995 p.67).

[3] *Relationship marketing* refers to all marketing activities directed toward establishing, developing, and maintaining successful relational exchanges (Morgan and Hunt, 1994 p.22).

customer segments. CRM unites the potential of relationship marketing strategies and information technology to create profitable, long-term relationships with customers and other key stakeholders. CRM provides enhanced opportunities to use data and information to both understand customers and co-create value with them. This requires a cross-functional integration of processes, people, operations, and marketing capabilities that is enabled through information, technology, and applications".

The trend that emphasises CRM as a business strategy has its root in the eighties, when Dwyer *et al*. (1987) highlighted the relationship aspect of buyer-seller behaviour instead of single transactions as the focus of marketing. Later, Reichheld and Sasser (1990) prove that companies focused on relationships can lead to significant advantages because customers tend to generate higher profits the longer they stay with the company. More recently, Richards and Jones (2008) classify some of the most common definitions of CRM into two related categories:

(i) CRM is often defined as a form of relationship strategy, for example: "*CRM is a comprehensive strategy and process of acquiring, retaining, and partnering with selective customers to create superior value for the company and the customer*" (Parvatiyar and Sheth, 2001 p. 5). Within this strategic view of CRM, firms can use CLV-CE estimations to develop customer acquisition and customer retention strategies (Blattberg *et al*., 2001b; Rust *et al*., 2000).

(ii) CRM is also often defined from a more operational view, for example: "*CRM allows companies to gather customer data swiftly, identify the most valuable customers over time, and increase customer loyalty by providing customised products and services*" (Rigby *et al*., 2002 p. 101). Therefore, within this operational view of CRM, the system facilitates the day-to-day interactions with customers (Van Bruggen and Wierenga, 2005), which can enable firms to use their customer databases and analytical tools to create opportunities for cross-selling new products and services to existing or new customers, i.e., to take decisions in the short-term.

From the strategic perspective, CRM is viewed as an asset (Srivastava *et al*., 2001), which is based on factors such as trust and reputation, is relatively rare and difficult for competitors to replicate, intangible, hard to measure and not nurtured. Additionally, relationships with customers are external assets to the firm, and therefore *available* to a firm, and *not owned*. Then, from the perspective of CRM, the task oriented to identify the most profitable customers (who will strengthen relationships in the long term, a priority for both academics and professionals in marketing), plays an important role in obtaining the firm competitive advantage. More specifically, central to the idea of CRM is the assumption that customers differ in their needs and in the value they generate to the firm. The way customers are managed should reflect these differences between them. In particular, from this strategic perspective, CRM is considered as an element to align business processes with customer strategies in order to increase customer loyalty and maximise profits over time (Rigby *et al*., 2002), or in other words, CRM pursues identifying profitable/valuable customers and then allocating the majority of resources and attention to these groups.

As firms increasingly see customers as important assets, methods for estimating CLV have been developed as an important strategic marketing tool (Ryals, 2002). In particular, the current interest that the marketing discipline is paying to the concept of CLV plays a crucial role in the CRM framework because CLV acts as an intermediary between marketing actions and decisions and the shareholders. CLV comprises a set of techniques that help companies to assess their portfolios of customers, improving CRM outputs and enabling marketing department to make their actions and decisions measurable. Using data, information, technology and applications, CLV allows companies to discover key customers and customer segments in order to understand them, develop long-term relationships with them and co-create value with them, the main goal of CRM (Payne and Frow, 2005 p. 168). Then, this CRM overall goal is aligned with CLV models' goal: CRM is therefore not about offering every single customer the best possible service, but about treating customers differently (i.e., according to their CLV). Concretely, the estimation of CLV is the key to managing customer relationships (Richards and Jones, 2008), because it is a measure to assess marketing decisions (Blattberg and Deighton, 1996) and to predict customer value of each customer in the database (Malthouse and Blattberg, 2005; Venkatesan and Kumar, 2004). This is essential as a base for a segmentation scheme and to decide about investment in (segments of) customers (Zeithaml *et al*., 2001)), and also to assess the whole customer base (Gupta *et al*., 2004) as a summation of the predicted CLV of all customers (this measure is called Customer Equity (CE)). A large group of researchers have recommended CLV for selecting customers and designing marketing programs (e.g., Reinartz and Kumar, 2003; Venkatesan and Kumar, 2004; Kim *et al*., 2006), because customers selected on the basis of CLV generate more profits than customers selected on the basis of other measures such as only socio-demographics (Reinartz and Kumar, 2003; Venkatesan and Kumar, 2004).

As a conclusion, it is interesting to remark that CRM is claimed to underpin theories on customer value (Mitussis *et al*., 2006), and therefore is inevitably linked with both CLV and CE (Weir, 2008 p. 808). Managers need to recognise that CRM is an enterprise wide concept that requires their businesses to identify opportunities to simultaneously enhance customer value while reducing costs, two effects that together create sustainable competitive advantage and result in greater short and long-term profitability (Bohling *et al*., 2006).

2.1.3. Customer Asset Management of Services (CUSAMS)

At this point it is interesting to remember that our main goal is to predict CLV (and CE) of a Spanish financial services retailer based on several core components considered essential in the customer-company relationship, in particular they are: **retention**, **product ownership**, **level of usage of products** and **contribution margin**. Customer management literature suggests that the components and drivers of CLV and CE fall into two categories: customer perceptions and customer behaviour (called unobservable and observable constructs respectively by Gupta and Zeithaml (2006)). We have focused on customer behaviour and have selected these variables because they reflect the three dimensions of a relationship: (1) length (retention), (2) breadth (cross-buying or product ownership), and (3) depth (increased usage/upgrading) (Bolton *et al*., 2004). It is well known that a multi-service provider generally depends on these core variables to increase the value of its customers (Wu *et al*., 2005).

Bolton *et al*. (2004) propose an integrated framework, called **CUSAMS (CUStomer Asset Management of Services)** that is a conceptual model about how marketing instruments influence purchase behaviours (see Figure 2). This framework reflects the length, depth, and breadth of customer-service provider relationships and thereby these three characteristics influence CLV (many other authors use these three components to configure a CLV model, such as Verhoef (2004) and Wu *et al*. (2005)). Therefore, CUSAMS framework enables service organizations to make a comprehensive assessment of the value of their customer assets (through length, breadth and depth of the relationship) and to understand the influence of marketing instruments on them (in particular price, service quality programs, direct marketing promotions, relationship marketing instruments (e.g., reward programs), advertising/communications, and distribution channels).

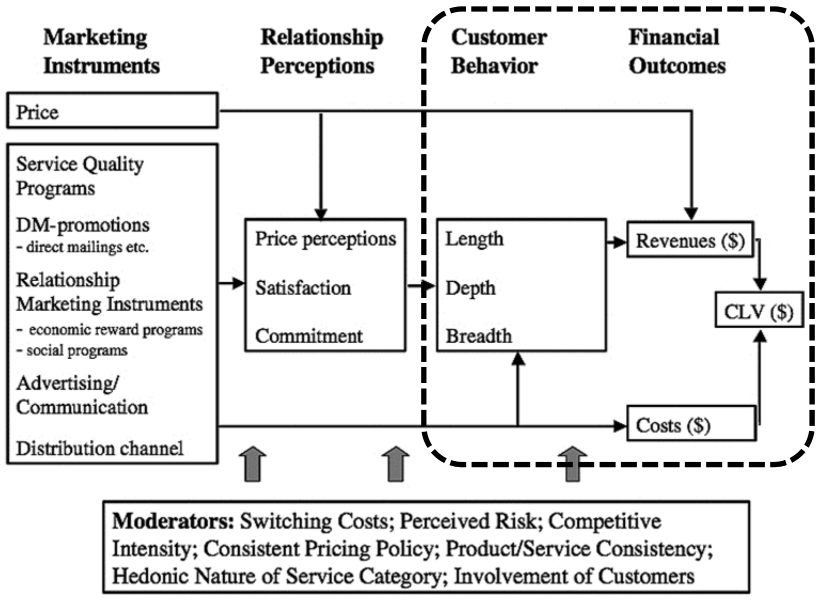
The foundation of the CUSAMS framework is a carefully specification of customer behaviours that reflect the length, depth, and breadth of the customer-service organization relationship (Verhoef, 2004), in particular:

(i) The *length or duration of a relationship* corresponds to customer retention (or defection), defined as the probability that a customer continues (or ends) the relationship with the organization.

(ii) The *breadth of the relationship* refers to the number of different products or services a customer is buying from a firm. It is reflected in cross-buying or 'add-on' buying. For instance, financial service providers may sell different financial services, such as loans, insurances and bank accounts. As a customer purchases more services, he or she will become more profitable.

(iii) The *depth of the relationship* refers to the purchase volume or the usage of a certain service or product. It is reflected in the frequency of product or service usage over time. It is also reflected in customers' decisions to upgrade and purchase premium (higher margin) products instead of low-cost variants.

Figure 2. CUSAMS framework (*)



NOTE: CUSAMS = Customer Asset Management of Services; DM = direct marketing; CLV = customer lifetime value.

(*) Source: Bolton *et al*. (2004)

2.2. Customer Lifetime Value (CLV)

Customer Lifetime Value has been studied under the name of *Lifetime Value (LTV)*, *Customer Equity (CE)*, *Net Present Value (NPV)*, *Customer Profitability (CP)*, or simply *Customer Value (CV)*. The differences among the definitions are slight (Hwang *et al*., 2004), and we explain the most important ones in the introduction of this research. In summary, Customer Lifetime Value (CLV), just as the name shows, assesses the long-term value of the relationship between customers and the company (Wu *et al*., 2005).

CLV was firstly defined by Kotler (1974 p.24) as the present value of the future profit stream expected over a given time horizon of transacting with the customer. More recently, CLV is defined as the present value of the future cash flows associated with a customer (Pfeifer *et al*., 2005). It is also formally defined as the sum of the discounted cash flows that an individual or a segment/group of individuals generates during his/her relationship with the company (Berger and Nasr, 1998), in other words, it is the net present value of benefits associated with each customer, once he or she has been acquired, after subtracting incremental costs associated with each customer (e.g., marketing, selling, production and service), over his or her entire life time with the company (Dwyer, 1997; Blattberg *et al*., 2008). In general, the CLV framework measures how changes in customer behaviour (e.g., increased purchase, customer retention or loyalty) could influence customers' future profits, or their profitability to the firm (Zhang *et al*., 2010), making a bridge between marketing and finance. In Table 5 several definitions of CLV are shown.

Table 5. Definitions of CLV (Hwang el al., 2004)

Definition	Reference
The net present value of all future contributions to overhead and profit.	Roberts and Berger (1989)
The net present value of a future stream of contributions to overheads and profit expected from the customer.	Jackson (1994)
The net present value of all future contributions to profit and overhead expected from the customer.	Courtheoux (1995)
The total discounted net profit that a customer generates during her life on the house list.	Bitran and Mondschein (1996)
Expected profits from customers, exclusive of costs related to customer management.	Blattberg and Deighton (1996)
The net present value of the stream of contributions to profit that result from customer transactions and contacts with the company.	Pearson (1996)
The net profit or loss to the firm from a customer over the entire life of transactions of that customer with the firm.	Berger and Nasr (1998)
The present value of all future profits generated from a customer.	Gupta and Lehmann (2003)

CLV and, by extension, CE are mainly based on the principles of contemporary finance of assets' valuation, more precisely the discounted cash flow (DCF) method, proposed by Rappaport in 1986, with two key differences (Gupta *et al*., 2006; Gupta and Lehmann, 2008):

(1) CLV is typically defined and estimated at an individual customer or segment level, allowing differentiation between customers based on profitability in order to identify customers who are more profitable than others and target them appropriately.

(2) Unlike in financial evaluations (e.g., Noone and Griffin, 1997; Smith and Dikolli, 1995; Van Raaij *et al*., 2003), CLV explicitly incorporates the possibility for future customer defection, typically through a retention rate.

Some researchers argue that CLV is based on the difference between customer revenues and customer costs (e.g., Calciu and Salerno, 2002; Gurau and Ranchhod, 2002; Mulhern, 1999), while others propose the contribution margin (e.g., Berger and Nasr, 1998; Malthouse and Blattberg, 2005; Reinartz and Kumar, 2000). Nevertheless, according to the financial theory, the value of any asset is the present value of its cash flows over time (i.e., cash inflows minus cash outflows). Few researchers have accurately applied the cash flow concept on CLV (e.g., Buhl and Heinrich, 2008).

Kumar and George (2007) explain that the value a customer brings to the firm is not limited to the profit from each transaction and is the total profit he/she may provide over the duration of his/her relationship with the firm. CLV is a concept that is forward looking, and the right definition and modelling should consider the essence of the concept as against rigid definitions (Jain and Singh, 2002). To be true to the notion of CLV, measures should look to the future, not to the past (Fader *et al*., 2005b), although unfortunately because of the challenges associated with forecasting future revenue streams, most empirical research of lifetime value has actually computed customer profitability solely on the basis of customer's prior behaviour. Finally, it is also important to note that CLV calculation helps companies to order customers according to their contribution to profits, which allows them to treat differently each one (Kumar and Rajan, 2009).

Considering the definitions above, our context and the data available, we define CLV as following:

CLV is the present value of a customer's current and future purchases. More specifically, ***CLV is the sum of the current and future contribution margins from the customers of the company, which depends on length, depth and breadth of their relationships, over their lifetimes of operation with company, taking into account the time value of money using a discount rate to adjust back the predictions about the future to the present.***

A comparison of different CLV models shows that, while an emphasis on retention is a common feature across them, conceptual differences in terms of accounting for existing customers and prospects, acquisition and the projection period exist. We give more details about these distinctive features of each model in next sections.

2.2.1. Individual versus aggregated CLV formulas

Several authors have developed different models to calculate CLV, some of them in an aggregate way (an average CLV for segments or the entire customer base), **and others in a disaggregate way** (individual CLV's). For example, **Jain and Singh (2002)** describe a basic model to calculate the average CLV of the firm, which considers only the current customers of the company, ignoring the past and potential future customers, the acquisition costs and other factors related stochastic purchase process and the timing of cash flow. The formula for this basic model is below, where i is the time period of the calculation (from a total of n periods of projected life of the customer under consideration), R_i is customer revenue in period i, C_i is the total cost incurred to generate revenues R_i in period i, and d is the discount rate. Therefore, the numerator is the net profit that has been obtained at each period, while the denominator transforms the net profit value into the current value.

$$CLV = \sum_{i=1}^{n} \frac{(R_i - C_i)}{(1+d)^{i-0,5}}$$

Another model to calculate the average CLV of the firm is developed by **Berger and Nasr (1998)**. These authors introduce a basic CLV model for a finite time period, based on three main assumptions: sales take place once a year, annual retention investment which is expected to occur at the middle of the purchase cycle (*M*) and both the retention ratio (*r*) as the gross margin contribution (*GC*) are assumed to be constant. Under these assumptions, the CLV is calculated as follows, where *n* is the length in years and *d* is the annual discount rate.

$$CLV = \{GC * \sum_{i=0}^{n} [r^i/(1+d)^i]\} - \{M * \sum_{i=1}^{n} [r^{i-1}/(1+d)^{i-0,5}]\}$$

When authors relax the assumption about the constant retention rate, gross contribution margin and promotional expenses, and assume that the purchase cycles can be longer or shorter than one year, the CLV equation gets modified as we show below, where $\pi(t)$ is the profit per customer in year *t*, which can be estimated separately using the appropriate equation for the profit curve.

$$CLV = \sum_{t=0}^{n} \pi(t) * \left[\frac{r^t}{(1+d)^t}\right]$$

Gupta and Lehmann (2003) and **Gupta *et al.* (2004)** also base their research on an aggregated CLV model. They consider an infinite time period in the calculation. For their first formula, the assumptions of the model are constant average margins (*m*) and constant retention rate (*r*), *i* is the discount rate.

$$CLV = \sum_{t=1}^{\infty} \frac{m * r^t}{(1+i)^t} = m\left(\frac{r}{1+i-r}\right),$$

$$where \left(\frac{r}{1+i-r}\right) is\ called\ margin\ multiple$$

Gupta and Lehmann (2003) and Gupta *et al*. (2004) simplify the last CLV formula developing two different cases, they are: (a) when margin growth occurs at a constant rate per period (g) and (b) when the average margins are constant.

$$(a)\ CLV = m\left(\frac{r}{1+i-r(1+g)}\right),$$

$$where\ \left(\frac{r}{1+i-r(1+g)}\right) is\ called\ margin\ multiple\ with\ margin\ growth\ g$$

$$(b)\ CLV = m\left(\frac{r}{1+i-r}\right),$$

$$where\ \left(\frac{r}{1+i-r}\right)\ is\ called\ margin\ multiple\ when\ margins\ are\ constant$$

Rust *et al*. (2000, 2004a) suggest another model to calculate an average CLV taking into account acquisition and retention of customers in the context of brand switching. For this reason this model incorporates the change matrix Markov process and requires information from the brand studied and the competitors. The sample provides information through a survey about the brand acquired in the last purchase, purchase probability of different brands and individual CE ratio (e.g., quality, price, convenience). The formula for this model is as follows, where T_{ij} is the number of purchases customer i is expected to make before firm j's time horizon, d_j is the discount rate of firm j, f_i is the average number of purchases customer i made in a unit of time (e.g., a year), V_{ijt} is expected purchase volume customer i brand j at time t, π_{ijt} is the contribution margin per unit of the brand j bought by consumer i at time t, B_{ijt} is the probability customer i buys brand j in time t (calculated using a Markov switching matrix). Rust *et al*. (2000, 2004a) also calculated CE (see definition and different formulas to calculate CE in the next section about CE concept).

$$CLV_{ij} = \sum_{t=0}^{T_{ij}} \frac{1}{\left(1+d_j\right)^{t/f_i}} * V_{ijt} * \pi_{ijt} * B_{ijt}$$

Other authors decide to take a further step in modelling the CLV and they develop CLV models at an individual or disaggregate level, such is the case of **Pfeifer and Carraway (2000)**. These authors mix RFM models with the CLV concept. RFM models traditionally use recency, frequency and monetary value of past customer responses to predict only the customer's short-term behaviour (for more details about this triad see Chapter 4, section 4.1, where we explain behavioural and attitudinal loyalty concepts). However, the CLV concept is related to model predictions in the long-term about customer value, but with very restrictive assumptions regarding customer behaviour, or do not provide a method for calculating the parameters of the model. Mixing these two approaches through Markov Chain Models (MCM), these authors develop a model that overcomes the shortcomings of both RFM and CLV approaches. Pfeifer and Carraway (2000) use RFM variables as a foundation of MC states (an MCM is a finite state machine with a probability measure assigned to each transition), where each state is defined from a different set of possible values of these variables. They define two main elements for this model: probabilities of transition (*P*) and reward vector (*R*). In particular, *P* between states reflects the probability that future customer behaviour will cause the customer to migrate from one state to the other, and *R* estimates the contribution and costs of each state. *R* is defined as we show below, where *NC* is the net contribution, i.e., the average contribution of the customer in each purchase, and *M* are marketing expenditures, i.e., the average marketing expense per customer over a defined time period.

$$R = \begin{Bmatrix} NC - M & if\ r = 1 \\ -M & if\ 1 < r < \max\_value \\ 0 & if\ r = \max\_value \end{Bmatrix}$$

The MCM theory provides two basic equations to calculate the value of each state (the first one is formulated for a finite time horizon and the second one for an infinite time period). These formulas take into account the time value of money by discounting from the future value to present value using the discount factor *d*. Each equation provides a vector (*V*) that defines the CLV of each state. *I* is the identity matrix.

$$V^T = \sum_{t=0}^{T} [(1+d)^{-1} * P]^t * R$$

$$\mathrm{V} = \lim_{T\to\infty} V^T = \{I - (1+d)^{-1} * P\}^{-1} * R$$

Also at an individual level (index *i* refers each customer), **Venkatesan and Kumar (2004)** want to identify the highest levels of consumer response to marketing communications across different channels in order to achieve optimal resource allocation between channels (giving priority to the most effective). We enclose their formula below, where $CM_{i,y}$ is the predicted margin contribution to customer i in purchase occasion y, r is the discount rate for money, $c_{i,m,l}$ are unit marketing costs for customer i channel m in year l, $x_{i,m,l}$ is the number of marketing contacts to customer i in channel m in year l, $frequency_i$ is the predicted purchase frequency for customer i, n is the number of years to forecast, and T_i is the predicted number of purchases made by consumer i until the end of the planning period.

$$CLV_i = \sum_{y=1}^{T_i} \frac{CM_{i,y}}{(1+r)^{y/frecuency_i}} - \sum_{l=1}^{n} \frac{\sum_m c_{i,m,l} * x_{i,m,l}}{(1+r)^{l-1}}$$

Following these authors, from the above equation it can be seen that CLV consists of three main components: (i) purchase frequency, (ii) contribution margin and (iii) marketing costs. For accurate measurement of CLV, researchers have to estimate these three components using suitable models and then combine these predictions to arrive at a single value representing the lifetime value of the customer (in monetary terms). Finally, all individual CLV's can be added to obtain the value of the company CE (see definition and different formulas to calculate CE in the next section about CE concept).

Berger *et al.* (2006) also develop a basic model to calculate CLV, where they do not take into account retention rate neither acquisition cost. In its more basic form, CLV is a function of a customer's future gross profits (revenue after deducting cost of goods sold and other marginal/variable costs). Future costs refer to those that are charged to individual customers, e.g., cost of services. The formula is enclosed below:

$$CLV_i = \sum_{i=1}^{n} \frac{(Future\ gross\ profits_{it} - Future\ costs_{it})}{(1+d)^t}$$

2.2.2. Deterministic versus stochastic CLV formulas

Another important distinction between CLV models is whether they are deterministic or stochastic (for more details see Chapter 3, section 3.1, about CLV and CE models classification). The main advantage of a stochastic CLV model is that it brings much more precision to CLV calculations by considering customer heterogeneity and building a probability model based on a sound behavioural theory.

A popular method to measure CLV in a non-contractual context is the ***negative binomial distribution (NBD)-Pareto model*** developed by the pioneers Schmittlein *et al.* (1987). In this model, past customer purchase behaviour (measure of purchase frequency and amount spent during a purchase) is used to predict the future probability of a customer remaining in business with the firm (the probability of each customer being alive). This probability can be used to estimate CLV (Reinartz and Kumar, 2000, 2003; Schmittlein and Peterson 1994). The NBD-Pareto model is applied in instances where customer lifetimes are not known with certainty (i.e., it is not known when a customer stops doing business with a firm). The model assumes that individual customer lifetimes with the firm are exponentially distributed. However, as was discussed by Schmittlein and Peterson (1994), in contexts (such as ours) where customer lifetimes are observed, the NBD-Pareto model has limitations and is not suitable. Other techniques to estimate CLV emerged, however NBD-Pareto model was not forgotten and other authors improved it later. Thus, we are going to return to this topic in the following paragraphs.

Colombo and Jiang (1999) were also pioneers proposing a stochastic way to model customer behaviour. This model uses observations of past responses to predict future responses, depending on different scenarios to carry out the task about which customers in the firm database to target with an offer (better offers to customers with higher expected contribution margin). Colombo and Jiang (1999) characterise each customer's buying behaviour with two probability distributions, one for the probability of purchase and one for the monetary amount the customer spends on an individual purchase. These authors do not assume that any of the variables were normally distributed. The total contribution to the firm under this policy may be considered as an average CLV for the set of customers that were targeted with the offer, although they do not calculate CLV explicitly and do not review the concept. One year later, **Pfeifer and Carraway (2000)** explicitly mix RFM variables (i.e., past customer behaviour) and CLV concept, as we have noted before (in section 2.2.1). Pfeifer and Carraway (2000) propose that a general class of mathematical models called Markov Chain Models (MCM) are appropriate for modelling customer relationships in a flexible way, because they can address the situations depicted in models proposed by Berger and Nasr (1998), Blattberg and Deighton (1996) and Dwyer (1997), i.e., retention and migration. In other words, MCM can be used to model both customer retention and customer migration situations. Additionally, the probabilistic nature of MCM allows to accounting for the inherent stochasticity in customer relationships and Pfeifer and Carraway demonstrate their use in various situations. These authors show how to get the key elements of MCM: the transition probability matrix and the reward vector (for more details see previous section 2.2.1).

Recently, some authors have done research into the same line as Pfeifer and Carraway (2000). For example, **Khajvand *et al.* (2011)** add a cross-selling variable to the famous RFM triad. This cross-selling variable is called *count item* (and collects the variety or number of products purchased by the customer). These authors segment customers in the company database by cluster analysis (using K-means Algorithm) and calculate the CLV for each segment. Other recent authors (e.g., **Etzion *et al.*, 2005; Hui-min *et al.*, 2006; Paauwe *et al.*, 2007**), have also based their research on this famous model (i.e., Pfeifer and Carraway, 2000). These authors amplify the information contained in each RFM triad, using variables about e-commerce, such as R_{buy} (time since last purchase), R_s (time elapsed since the last visit to the Web site of the company), F_{buy} (total number of acquired products) or F_s (total number of logins on the website of the company), and even cross-buying (measured as the total number of products purchased by the customer), among other variables. The philosophy of Khajvand *et al.* (2011) is similar to what we want to implement in our research (adding more information to the simplest RFM triad), although we are going to use a wider set of variables to define more specifically all of the predictors of retention, product ownership, product usage and contribution margin (for more details see Chapter 4). These variables allow us to obtain a more complete customer value model and to discover which variables CLV depends on. We also want to implement the model through more powerful data analysis methodologies: predictive and probabilistic.

Despite the advantages of MCM, there are some critical assumptions underlying this method. Specifically, in MC models time period of purchase by all customers is assumed to be the same, and fixed. The calculation of transition probabilities is critical to the success of such models and these probabilities are not easy to estimate. In addition, MCM do not account for heterogeneity in the underlying behaviour characteristics, which can lead to misleading interferences about the nature of buying behaviour. To overcome these disadvantages, **Fader *et al.* (2005b)** present a new model that also links the well-known RFM paradigm with CLV. Fader *et al.* (2005b) use Bayes' theorem to develop their model, based on the premise that observed behaviour is a realization of latent traits. In particular, these authors use this theorem to estimate a person's latent traits as a function of observed behaviour and then predict future behaviour as a function of these latent traits.

At the heart of the Fader *et al.*'s (2005b) model is the previously mentioned Pareto/NBD framework for the flow of transactions over time in a non-contractual setting (proposed by Schmittlein *et al*. in 1987[4]). An important characteristic of their model is that RFM measures are enough statistics for an individual customer's purchasing history. That is, rather than including RFM variables in a scoring model simply because of their predictive performance as explanatory variables, Fader *et al*. (2005b) formally link the observed measures to the latent traits and show that no other information about customer behaviour is required to implement the model. Note that this model implicitly assumes a constant retention rate (exponential dropout rate). Further, this model does not typically incorporate marketing covariates, and its focus is to simply predict the probability of a customer being alive rather than identify which factors influence retention. Finally, this model assumes Poisson transaction rates, which are not well suited for situations where customers have non-random or periodic purchase behaviour (e.g., grocery shopping every week). Nonetheless, it provides a good benchmark.

In line with Fader *et al*. (2005b), other researchers have used probability models to perform forward-looking customer-base analysis using the framework that states that observed behaviour is the outcome of an underlying stochastic process. Fader and Hardie (2009) explain in a very clear way how to develop a model of this kind:

> *"The starting point is to specify a mathematical model in which the observed behaviour is a function of an individual's latent behavioural characteristics (i.e., past = f (θ)). This is done by reflecting on what simple probability distribution (e.g., Poisson, binomial, exponential) can be used to characterize the observed behaviour (in many cases, including those to be discussed in this paper, observed behaviour may be characterized using a combination of these basic probability distributions). By definition, we do not observe an individual's latent characteristics (θ). Therefore, the next step is to make an assumption as to how these characteristics vary across the customer base by specifying a mixing distribution that captures the cross-sectional heterogeneity in θ (the choice of distribution(s) is typically*

[4] The Pareto/NBD model is based on six assumptions: (i) customers go through two stages in their 'lifetime' with a specific firm: they are 'alive' for some period of time, and then become permanently inactive; (ii) while alive, the number of transactions made by a customer follows a Poisson process with transaction rate λ; (iii) a customer's unobserved 'lifetime' of length τ (after which he is viewed as being inactive) is exponentially distributed with dropout rate μ); (iv) heterogeneity in transition rates across customers follow a gamma distribution with shape parameter r and scale parameter α; (v) heterogeneity in dropout rates across customers follows a gamma distribution with shape parameter s and scale parameter β; (vi) the transaction rate λ and the dropout rate μ vary independently across customers. The population parameters r, α, s and β are unknown so they must be estimated.

driven by the dual criteria of flexibility and mathematical convenience). Combining this assumption with the distribution for individual-level behaviour gives us a mixture model, which characterizes the behaviour of a randomly chosen customer. After fitting the mixture model to the data, a straightforward application of Bayes' theorem enables us to make inferences about an individual's latent characteristics (θ) given his observed behaviour. We can then make predictions regarding future behaviour as a function of the inferred latent characteristics. Note that there is no attempt to explain the variation in θ as a function of covariates; we are, in most cases, content to capture the variation using probability distributions alone. This two-step approach (θ = f (past) and future = f (θ)) can be contrasted with the single-step approach (future = f (past)) associated with the use of regression models (and more sophisticated data mining procedures)" (Fader and Hardie, 2009).

A key assumption of the model developed by Fader *et al*. (2005b) is the independency between the number of transactions of a customer and the related profit per transaction. Fader *et al*. (2005b) propose natural extensions to their model for non-contractual settings: firstly, introducing marketing mix variables; secondly, relaxing the assumption of independence between the distribution of monetary value and the underlying transaction process (this could be accommodated by replacing their respective gamma distributions with a bivariate Sarmanov distribution that has gamma marginal); thirdly, relaxing the assumption of constant contribution margin per transaction; and finally, running the model across multiple cohorts to obtain an accurate picture of the value of an entire customer base. **Glady *et al*. (2009b)** take advantage of Fader *et al*.'s (2005b) proposed extensions to modify their model, not relying on the independence assumption. The purpose of their paper is to modify Pareto/NBD-based approach for CLV prediction, using a data set provided by a Belgian financial service institution. Glady *et al*. (2009b) show that the newly proposed method has better forecasting performance than the traditional Pareto/NBD model, and it also outperforms a standard regression approach. Glady *et al*. (2009b) only consider transactional data in a banking context and propose the inclusion of socio-demographic explanatory variables for the CLV prediction, for example as regressors when studying the dependence between the number of transactions and the average profit per transaction.

Another approach that can naturally incorporate past behavioural outcomes into future expectations is a ***Bayesian approach*** (Rossi and Allenby, 2003). Bayesian methods can incorporate such prior information in the structure of the model easily through the priors of the distributions of the drivers of CLV (Abe, 2009b; Borle *et al*, 2008). Furthermore, this approach can be used in any context (that is, contractual or non-contractual). In particular, Borle *et al*. (2008) and Abe (2009b) call for the inclusion of a rich set of covariates in their Hierarchical Bayesian framework to estimate CLV. Our variables could enrich the estimation of CLV through Hierarchical Bayesian approach and at the same time we could prove if these variables have a real effect on the value of each customer.

At this point it is interesting to mention several works that use ***Hierarchical Bayesian approach (HB)*** to develop customer valuations in a more flexible and easy way. Mathematically, the approach pursued in a Pareto/NBD model is so-called *empirical Bayes*, whereby the same data are used for the likelihood (customer specific purchase and survival functions) as well as for estimating the prior (a mixture distribution), resulting in the overestimation of precision. *Empirical Bayes is an approximation of a hierarchical Bayesian method in the Bayesian paradigm* and about the *hierarchical Bayesian method* no threat is posed if the sample size is large or the mixture distribution is estimated from separate data (Gelman *et al*., 1995).

One of the studies that use Hierarchical Bayesian approach to develop a customer value model is performed by **Borle *et al.* (2008)**. These authors propose a hierarchical Bayesian model that works better than both the Pareto/NBD and the RFM models in a special and 'new' context called '*semi-contractual setting*' (membership based direct marketing company) in predicting CLV. This approach can naturally incorporate past behavioural outcomes (prior information) into future expectations easily through the priors of the distributions of the drivers of CLV (Rossi and Allenby, 2003). Furthermore, this approach can be used in any context. The model jointly predicts a customer's risk of defection and spending pattern at each purchase occasion. This information is then used to estimate the lifetime value of each customer of the firm at every purchase occasion. One potential drawback of this analysis may be the availability of the appropriate covariates, and the authors propose as future research stream to incorporate a richer set of these variables, as we propose in our research. Additionally, **Abe (2009b)** also uses Hierarchical Bayesian framework to extend a Pareto/NBD model of customer-base analysis. In particular, he proposes for further research to incorporate the effect of marketing activities as the covariates for frequency dropout and spending. He also explains in a very clear way what are the ***advantages of Bayesian Hierarchical approach to model CLV***, and we refer them here:

(1) The HB model is more flexible because it implements dependence assumption of purchase rate, dropout rate and spending process parameters. The independence between them is a critical assumption in a Pareto/NBD model that can be relaxed with HB. Managerially, this assumption restricts that shopping frequency, lifetime, and spending per trip are not related. The parameter estimation of a Pareto/NBD model might be biased if this independence assumption is violated. HB model not only accommodates correlated data, but also allows performing statistical inference of the independence assumption on data.

(2) The HB model, whereby customer-specific parameters are a function of covariates, can be constructed and estimated with ease. Substantively, such HB models can shed light on interesting yet conflicting findings in CRM. For example: What are the characteristics of loyal customers (with long lifetime)? and do loyal customers spend more? Previous research studied such issues with a two-step approach: lifetime duration is firstly estimated to identify loyal customers, and then customer characteristics (explanatory variables) are related to the lifetime duration (that is the dependent variable). The HB model, whose dropout parameter is a function of customer characteristics, can be estimated in one step, providing the correct measure of error for statistical inference.

(3) Because the distribution of the purchase rate, dropout rate, and spending parameter are estimated at the individual level as by-product using the MCMC method, the distribution of any customer-specific statistics, such as mean and variance, can be computed by simple algebra. Such statistics of managerial relevance include a probability of being active at a certain point in time, an expected lifetime, a one-year survival rate, and an expected number of transactions in a future period. The fact that a distribution rather than a point estimate of a statistic is obtained also permits the application of statistical inference at the individual level without being restricted by the asymptotic properties.

Only few studies have compared the performance of complex versus noncomplex models for customer purchase behaviour and CLV prediction. One of them is **Wübben and von Wangenheim's (2008)** work, where they show that a model does not necessarily have to be sophisticated in order to precisely forecast a customer's transactions, especially with respect to managerial relevance and applicability. Wübben and von Wangenheim (2008) prove that simple heuristics using initial and repeated purchase data perform at least as well on the individual level as the stochastic models as Pareto/NBD and BG/NBD. Similarly, **Donkers *et al.* (2007)** compare a set of models including a status quo model, a Tobit II model, univariate and multivariate choice models, and duration models (survival analysis). These authors find that the simple models perform well, and using complex methods instead of simple models for CLV prediction in a contractual setting (for instance, an insurance company) does not substantially improve the predictive accuracy. **Zhang *et al.* (2010)** also find this same rational, that is, more mathematically sophisticated models (e.g., NBD model) do not substantially outperform less-sophisticated models (e.g., RFM model). In conclusion, this fact is meaningful for marketing researchers who seek research simplicity or attempt to avoid computation intensity.

All these examples are only a small sample of the wide range of models to measure CLV, although despite this wide range of models developed to measure the value of customers throughout their life cycle with a particular company (that is, using the CLV measure), there is no consensus about the best method for their calculation (Jain and Singh, 2002), and it depends on the context of application. Moreover, the appropriate way to model CLV depends on the nature of the business and on the aims of the modelling (Crowder *et al*., 2007). In Chapter 3 (about CLV and CE models classification), we show several criteria used to classify CLV models in a more complete and clearer way.

2.3. Customer Equity (CE)

The long-term value of a firm is largely determined by the value of the customer relationships of this firm, which result in the **Customer Equity (CE)** (Aravindakshan *et al*., 2004). Then, the concepts of **Customer Lifetime Value** and **Customer Equity** are related and sometimes are considered equivalent in the literature. While there is a general agreement on the definition of the first (i.e., CLV), there are different definitions of CE.

Some authors define CE as the average CLV less acquisition cost (Berger and Nasr, 1998; Blattberg and Deighton, 1996; Blattberg *et al*., 2001a). In particular, **Berger and Nasr-Bechwati (2001)**, who developed an aggregate-level approach of CE for balancing acquisition and retention expenditures, explain that the difference between CE and CLV is that CE takes acquisition cost into consideration. We show this aggregate-level approach of CE in the following formula, where a is the acquisition rate (proportion of solicited prospects acquired), given a specific level of acquisition costs (A), m is the margin (in monetary units) on a transaction, A is the acquisition cost per customer, R is the retention cost per customer per year, r is the yearly retention rate, and d is the yearly discount rate (appropriate for marketing investments):

$$CE = am - A + a * \left(m - \frac{R}{r}\right) * \left[\frac{r^n}{1 - r^n}\right], with\, r^n = r/(1 + d)$$

Blattberg *et al.* (2001a) develop another formula to calculate CE at a customer segment or cohort level for a finite time period. CE is calculated as the sum of return on acquisition, return on retention and return on add-on selling across an entire customer portfolio of the firm. One component of the equation computes returns from acquisition as the contribution from newly acquired customers minus the cost of acquiring them. The other component of the equation calculates the expected profits from future sales made by these newly acquired customers adjusted for retention rate and time value of money. Thus, CE (t) is the value of CE for newly acquired customers in the time period *t*, where $N_{i,t}$ is the number of potential consumers for the segment *i* at time period *t*, $a_{i,t}$ is the acquisition probability in period *t* for a customer in segment *i*, $\rho_{i,t}$ is the retention probability in period *t* for a customer in segment *i*, $B_{i,a,t}$ are marketing cost per prospect (*N*) for acquiring customers at time *t* for segment *i*, $B_{i,r,t}$ are the marketing costs in time period *t* for retained customers for segment *i*, $B_{i,AO,t}$ are the marketing costs of add-on selling in period *t* for segment *i*, *d* is the discount rate, $S_{i,t}$ are the sales of products and services offered by the firm at time *t* for segment *i*, $c_{i,t}$ are the cost of goods in period *t* for segment *i*, the number of segments are indexed by *i* and the number of periods are indexed by *k*. The formula to get CE is:

$$CE(t) = \sum_{i=0}^{I} [N_{i,t} * \alpha_{i,t} * (S_{i,t} - c_{i,t}) - N_{i,t} * B_{i,a,t}$$
$$+ \sum_{k=1}^{\infty} N_{i,t} * \alpha_{i,t} \left(\prod_{j=1}^{k} \rho_{j,t+k} \right) * (S_{i,t+k} - c_{i,t+k} - B_{i,r,t+k} - B_{i,AO,t+k})$$
$$* (\frac{1}{1+d})^k]$$

Other authors propose that the CE of the firm is formed by the CLV's of all the current and potential customers (**Zhang *et al.*, 2010**), which has been discovered to be a good proxy measure of the equity-market valuation of the firm (Gupta *et al.*, 2004). Compared to CLV, CE is a macro-level measure that can be applied directly to understand equity market reactions to marketing actions. In particular, CE is defined as the average value of the entire database of customers or customer segments (**Wiesel and Skiera, 2005**), or in other words it is the customer value at the firm level (**Kumar and Shah, 2009**). For this research, we use this second definition of CE. In the formula enclosed below, CLV_i is customer lifetime value of each customer *i* and *N* is the total number of customers, which includes the current customer base (or each of the segments) and future customers:

$$CE = \sum_{i=1}^{N} CLV_i$$

Other similar way to get CE suitable for panel data (where the effect of competition is collected), is the following formula (**Rust *et al.*, 2000, 2004a**), where $mean_i\ (CLV_{ij})$ is the average lifetime value for firm *j*'s customers *i* across the sample (they do not estimate individual CLV's) and *POP* is the total number of customer in the market across all brands (effect of competition is included):

$$CE_j = mean_i\left(CLV_{ij}\right) * POP$$

We can also make a distinction between different CE concepts: (i) *static CE*, (ii) *dynamic CE*, (iii) *CE contribution* and (iv) *CE elasticity* (Villanueva and Hanssens, 2007). The different definitions are below:

(i) *Static CE* is the present value (not future value) generated by all the individual customers or cohorts of customers along their relationship with the firm (Yoo and Hanssens, 2005).

(ii) *Dynamic CE* is defined as the discounted sum of both current and future cohorts' CE. Because its customers ultimately generate most of the cash flow of a company, it has been suggested that this measure is a good proxy for the value of a firm because it accounts for both current and future relationships (Gupta *et al*., 2004). In particular, Gupta *et al*. (2004) calculated the correlation between CE and Market Value, proving that the CE is less volatile than Market Value. CE is a useful measure for companies interested in obtaining equilibrium on its long-term strategies (Drèze and Bonfrer, 2003).

(iii) *CE contribution*. An idea behind the CE management is that the CE measure can provide a basis for calculating the R.O.I. of any investment (e.g., newly acquired customers by a specific promotion) (Weir, 2008). Rust *et al*. (2004a) show the formula for this measure, where ΔCE is the improvement in CE produced by the expenditures and *E* the discounted expenditure stream:

$$R.O.I. = (\Delta CE - CE)/E$$

(iv) *CE elasticity,* which measures the increased percentage in CE from a 1% change in the marketing mix (e.g., advertising spending) or in any parameter of the CE specification (e.g., retention rate) (Yoo and Hanssens, 2005).

Considering the definitions above, our context and the data available, we define our CE in the same line of Zhang *et al*. (2010):

CE is formed by the CLV's of all the current and potential customers of the firm, which has been found to be a ***good proxy measure of the equity-market valuation of the firm*** (Gupta *et al*., 2004). In general, we refer to a ***dynamic CE***.

Once the analytical models to estimate the CLV and CE are determined, the drivers and components of CLV and CE can be identified and subsequently exploited, e.g., for specific marketing actions. In this sense, the CLV-CE management is configured as a dynamic and interactive financial valuation technique used to optimise marketing strategies (e.g., acquisition, retention, cross-selling) (Berger and Nasr, 1998).

Chapter 3. LITERATURE REVIEW ON CLV AND CE APPLICATIONS AND MODELS

3.1. CLV and CE models classification

We have developed a classification of CLV models by combining several criteria taken into account by previous researches about this topic. The majority of these articles were published in journals with an 'impact index' in accordance with the Social Science Citation Index (SSCI)[5] (e.g., *Harvard Business Review*, *Journal of Interactive Marketing*, *Journal of Marketing*, *Journal of Marketing Research*, *Journal of Service Research* and *Management Science*), which guarantee the quality of the studies. The objective of this classification is to corroborate the importance of CLV and CE models in marketing. Additionally, we want to offer a global and integral view of CLV-CE models that serves as a guide describing key requirements for developing these types of models. The criteria considered are: type of relationship between customer and company (section 3.1.1); whether the analysis is historical or predictive (section 3.1.2); whether the analysis is deterministic or stochastic (data analysis methodology) (section 3.1.3); source of data (section 3.1.4); whether the effect of competition is included (section 3.1.5); and the aggregation level of the data for CLV calculation (section 3.1.6). Firstly, we offer Table 6, to guide the reader easily into a deeper explanation enclosed below.

[5] Some of the articles within this collection are published in Journals without this 'impact factor' (e.g., *Decision Support Systems*, *European Journal of Operational Research*, *Journal of Consumer Marketing*, *Journal of Database Marketing* and *Journal of Relationship Marketing*). These exceptions were considered because these papers have received a significant number of cites from other articles about this topic.

Table 6. Summary of the CLV and CE models classification

Criteria	**Values**		**Examples**
3.2.1. Type of relationship between customer and company	(i) *lost for good/retention/contractual setting*		Blattberg and Deighton (1996); Wiesel *et al*. (2008)
	(ii) *always a share/migration/non-contractual setting*		Venkatesan and Kumar (2004); Rust *et al*. (2004a)
	(iii) *semi-contractual setting*		Borle *et al*. (2008)
3.2.2. Historical or predictive analysis?	(i) *historical CLV models*		Reinartz and Kumar (2003); Venkatesan and Kumar (2004)
	(ii) *predictive CLV models*		Gupta *et al*. (2004); Malthouse and Mulhern (2008)
3.2.3. Deterministic or stochastic analysis?	*Deterministic equations*	(i) *RFM models*	Kahan (1998); Marcus (1998); Miglautsch (2000)
		(ii) *Growth and diffusion models*	Gupta *et al*. (2004); Hogan *et al*. (2003)
	Stochastic process	(i) *Probability model*	Abe (2009b); Borle *et al*. (2008); Libai *et al*. (2002); Reinartz and Kumar (2000, 2003); Rust *et al*. (2004a)
		(ii) *Econometric models*	Van den Poel and Larivière (2004)
		(iii) *Persistence models*	Villanueva *et al*. (2008); Yoo and Hanssens (2005)
		(iv) *Computer science models*	Neslin *et al*. (2006)
3.2.4. Source of data	(i) *Database of customers*		Venkatesan and Kumar (2004); Verhoef and Donkers (2001)
	(ii) *Survey*		Rust *et al*. (2004a)
	(iii) *Public reports*		Gupta *et al*. (2004); Gupta and Lehmann (2003)
	(iv) *Panel data*		Yoo and Hanssens (2005)
	(v) *Managerial judgments*		Blattberg and Deighton (1996); Ryals (2005)
3.2.5. Is effect of competition included?	Yes		Reinartz *et al*. (2005); Yoo and Hanssens (2005)
	No		Villanueva *et al*. (2008); Ryals (2005)
3.2.6. Level of aggregation in the data for the CLV calculation	(i) *Calculation of average CLV from aggregate measures*		Blattberg and Deighton (1996); Gupta and Lehmann (2003); Gupta *et al*. (2004)
	(ii) *Calculation of individual CLV from individual measures*		Drèze and Bonfrer (2002); Reinartz and Kumar (2000, 2003); Venkatesan and Kumar (2004)

In the following paragraphs, we explain in detail the proposed CLV-CE models classification.

3.1.1. Type of relationship between customer and company

Traditionally, researches have considered two types of customer-company relationships to calculate CLV, depending on the way this relationship is interpreted: (i) *lost for good/retention/contractual setting* (Blattberg and Deighton, 1996; Wiesel *et al*., 2008) and (ii) *always a share/migration/non-contractual setting* (Venkatesan and Kumar, 2004; Rust *et al*., 2004a). In order to refer to these two different situations we have chosen the concepts suggested by Reinartz and Kumar (2000), i.e., contractual and non-contractual settings. Recently, another type of relationship has been termed as (iii) *semi-contractual* (Borle *et al*., 2008). The two traditional types of relationships have been named differently according to different authors, as we can see in Table 7.

Table 7. Type of relationship between customer and company

Authors	**(i) Contractual**	**(ii) Non-contractual**
Jackson (1985)	Lost for good	Always a share
Dwyer (1997)	Retention	Migration
Reinartz and Kumar (2000)	Contractual	Non-contractual

These two behaviours (i.e., contractual and non-contractual) display different patterns that should be considered in the implementation stage of the model. In particular, they imply to take a previous decision before starting to solve the problem (that is, to take the assumption of a contractual or a non-contractual setting), choosing a suitable scenario in order to apply the correct methodology. Some authors have noted that it is totally unacceptable to apply a model developed for a contractual setting in a non-contractual one and vice versa (Fader and Hardie, 2009). Despite this fact, Borle *et al*. (2008) and more recently Abe (2009a, 2009b) have applied Hierarchical Bayesian approach to calculate CLV. This methodology is more flexible and it can accommodate both situations (i) and (ii). In particular, Borle *et al*. (2008) point out that a third kind of relationship between customer-company is possibl, it is called (iii) *semi-contractual*.

The first case **(i) *lost for good/retention/contractual setting*** implies that customers have made long-term commitments to a vendor because switching vendors is costly and assets dedicated to the transaction cannot be redeployed easily (Dwyer, 1997). The time at which a customer becomes inactive (attrition) is observed by the firm because the company maintains a record of each customer by establishing contracts with them, therefore retention rate (or its opposite, churn rate) is a directly observable variable. This kind of relationship considers that a customer remains alive as long as he/she generates transactions. This means that if at some given moment customers do not renew their contracts or do not generate any transaction they can be considered as 'lost for good' or as 'ex-customers'. It also means that if an ex-customer buys again he/she is considered as a new customer and one deals with an acquisition rather than a customer retention issue. Dwyer (1997) stated that a lost for good situation is best modelled as customer retention problem.

According with this first category of models, Gupta and Lehmann (2008) propose several models to calculate customer retention (in a contractual setting), they are explained below.

(1) *Logit or Probit models*. In contractual settings where customer defection is observed, it is easy to develop a Logit or a Probit model of customer defection. Due to its easy estimation, this approach is usually used in the industry. This model takes the familiar Logit (or Probit) form as follows, where *X* are covariates:

$$P(Churn) = \frac{1}{1 + \exp(\beta X)}$$

(2) *Hazard models*. The inter-purchase time can also be modelled using a hazard model (Logit or Probit models are a form of discrete time hazard models). Hazard models fall into two broad groups: *accelerated failure time (AFT)* or *proportional hazard (PH) models*.

(2.1) *AFT models* have the following form (Kalbfleisch and Prentice, 1980), where *t* is the purchase duration for customer *j* and *X* are the covariates. If σ = 1 and μ has an extreme value distribution, then we get an exponential duration model with constant hazard rate.

$$\ln(t_j) = \beta_j X_j + \sigma \mu_j$$

Different specifications of σ and μ lead to different models, such as Weibull or generalised gamma usual to model relationship duration (e.g., Venkatesan and Kumar, 2004).

(2.2) *PH models* are another group of commonly used duration models (e.g., Reinartz and Kumar, 2003). These models specify the hazard rate (λ) as a function of baseline hazard rate (λ_0) and covariates (*X*).

$$\lambda(t; X) = \lambda_0(t)\exp(\beta x)$$

Different specifications for the baseline hazard rate provide different duration models, such as exponential, Weibull or Gompertz.

(3) *Computer Science models*. The marketing literature has typically favoured structured parametric models (e.g., Logit, Probit, hazard models). These models are based on utility theory and are also easy to interpret. In contrast, the vast computer science literature in data mining, machine learning and non-parametric statistics has generated many approaches that emphasise predictive ability. These include projection-pursuit models, neural network models, decision tree models, spline-based models (such as Generalised Additive Models (GAM) and Multivariate Adaptive Regression Splines (MARS)), and support vector machines. Many of these approaches may be more suitable to study customer churn where the researcher has a very large number of variables. The sparseness of the data in these situations inflates the variance of the estimates making traditional parametric and nonparametric models less useful.

Later, Fader and Hardie (2009) collect several probability models for customer-base analysis. For this first case (contractual setting), Fader and Hardie (2009) make distinctions between two possible kinds of transactions: continuous and discrete in time (this last kind of transactions can only be attended at discrete points in time or can occur at any point in time but are treated as discrete by management).

(1) The general probability model to model a *contractual setting for discrete-time duration data* (e.g., magazine subscriptions, insurance policy) is the shifted-beta-geometric sBG model (e.g., Fader and Hardie, 2007). These models assume that at the end of each contract period, an individual remains a customer of the firm with constant retention probability, and the differences in this constant retention probability are characterised by the beta distribution. The aggregate retention rate with the sBG model is an increasing function of time (increasing retention rate over time).

(2) On the other hand, the general probability model to model a *contractual setting for continuous-time data* (e.g., credit card, mobile phone usage) is the exponential gamma (or Pareto of the second kind) distribution (e.g., Morrison and Schmittlein, 1980). The continuous time analogous to the churn rate is the hazard function. The hazard function of

the exponential gamma model decreases over time, so the aggregate churn rate is a decreasing function of time.

The assessment of customers assuming this kind of relationship (that is, lost for good or contractual setting) will only take into account the customers' probability to remain active from one period to another, and this is the number of successive periods during which the customer is active. This scenario is questionable: it understates the CLV, because it does not allow a defected customer to return to the company (Rust *et al*., 2004a).

The second case **(ii)** ***always a share/migration/non-contractual setting*** implies that customers may rely on several vendors and can adjust their share of business done with each (Dwyer, 1997). The time at which a customer becomes inactive is unobserved by the firm, i.e., customers do not notify the firm "*when they stop being a customer. Instead they just silently attrite*" (Mason, 2003 p.55). This type of relationship considers that customers can reappear (turn up again) after some periods during which they did not make transactions. In other words, after a certain period of inactivity a customer can return to the company and he/she is not considered a lost customer. For this setting, Dwyer (1997) described a customer migration model, using purchase recency to predict purchase behaviour.

Therefore, in non-contractual settings firms have to infer whether or not a customer is still active. Most companies define an active customer based on simple rules-of-thumb (for example, eBay defines a customer to be active if he/she has bid, bought or listed on its site during the last 12 months). However, researchers prefer to base themselves on statistical models to assess the probability of retention. The importance of retention has led researchers to spend a large amount of time and energy in modelling this component of CLV. Broadly speaking, these models can be classified into several categories (Gupta and Lehmann, 2008):

(1) *Markov models*. While most previous models implicitly assume that a customer who defects is 'lost for ever', in Markov models customers are allowed to switch among competitors and therefore they are considered as 'always a share'. These models estimate the transition probabilities of a customer in a certain state moving to other states. Using these transition probabilities, CLV can be estimated, for example, as Pfeifer and Carraway's model (2000).

(2) *Probability or stochastic models*. They are a special type of retention hazard models that were firstly proposed by Schmittlein *et al*. (1987). These models use recency and

frequency of purchases to predict probability of a customer being alive in a specified future time period. We give more details about these models below.

According with the "*buy until you die*" framework (Schmittlein *et al*., 1987), a customer's relationship with a firm has two phases: (1) he/she is 'alive' for an unobserved period of time and (2) he/she becomes permanently inactive. Having this idea in mind, Fader and Hardie (2009) study this topic in detail to better differentiate between two possible kinds of transactions: continuous and discrete.

(1) In a *non-contractual setting with continuous transactions*, (a) the standard probability model used to model repeat-buying behaviour is NBD, and (b) the Pareto (gamma mixture of exponentials) to model the customer's unobserved lifetime after the customer is permanently inactive. The resulting model is called **Pareto/NBD** (e.g., Fader *et al*., 2005b; Glady *et al*., 2009b; Reinartz and Kumar, 2003). Some authors get to model the heterogeneity in customer's dropout (referred here as b) through a beta-geometric distribution (BG), resulting in a joint model called **BG/NBD** (e.g., Fader *et al*., 2005a). Other researchers have chosen other probability distributions to model inter-purchase time (e.g., Wu and Chen, 2000; Allenby *et al*., 1999), even other ways to capture non-stationarity in buying rates, but they are models more difficult to implement because generally require the full transaction history (i.e., cannot be estimated using only recency and frequency data). Furthermore, no one has yet derived expressions for quantities such as P(alive) and conditional expectations, which are central to any forward-looking customer-base analysis exercise.

(2) In a *non-contractual setting with discrete transactions*, (c) the standard probability model used to model the purchasing process is Bernoully (e.g., Colombo and Jiang, 1999; Fader *et al*., 2004). It assumes that the probability of making (or not) a purchase in one period is independent of the preceding period (to relax this assumption, some authors have adopted some type of first order Markov process (Pfeifer and Carraway, 2000; Rust *et al*., 2004a)). (d) The defection process is modelled by another Bernoully distribution. Therefore, the joint model is called **BG/BB**.

Pareto/NBD, BG/NBD and BG/BB models do not require information on when transactions occurred to predict future purchasing activity. The first two models only need recency (time between transactions) and frequency (of transactions) variables. The last model only requires a binary string, where 1 indicates a transaction and 0 otherwise.

As Calciu (2008 p. 223; 2009 p. 261) explains, the 'always a share' behaviour is the alternative scheme to 'lost for good' in what is known as a dichotomy. CLV here comes not only from surviving customers but also from customers allowed to reactivate after a given number of inactive periods. Customers are considered 'lost for good' only after exceeding that number of successive periods of inactivity. By reducing the tolerated number of successive periods of inactivity to zero, the 'always a share' model reduces to the 'lost for good' model. Therefore and as a conclusion, 'lost for good' is a special case of 'always a share', a more general and complete model (Rust *et al*., 2004a).

As an extension to the previously mentioned non-contractual setting, where firms cannot know when a customer becomes inactive, intuitively they can apply rules (conventions) based on the RFM amount of past purchases in order to decide whether or not a customer is still active. By fixing RFM states, based on past behaviour, transition probabilities from one state to another can be computed and organised into a matrix of transition probabilities in order to form a Markov Chain. A detailed discussion of the matrix approach applied to customer migration is found in the study by Pfeifer and Carraway (2000) (for more details about this research see previous section 2.2.1). Other researchers that use this matrix approach to calculate CLV are Bitran and Mondschein (1996) and Rust *et al*. (2004a), who define transition probabilities matrices between brands (effect of competition is included, see section 3.1.5) that they vary over time following a Logit model.

Finally, the third case is a **(iii)** ***semi-contractual setting*** (Borle *et al*., 2008). Borle *et al*. (2008) selected a special context for their study: membership-based direct marketing company; examples of such companies are membership-based clubs such as music clubs, book clubs, and other types of purchase-related clubs. This context has elements of both contractual and non-contractual settings, a scenario that has not been analysed in-depth previously (Singh and Jain, 2007). As in a contractual setting, the firm knows customer lifetime information of past customers with certainty (i.e., the time when a membership begins and the time when he/she ends are known once these events happen for each customer). On the other hand, as in a non-contractual setting both the purchase timing and spending on purchases do not happen continuously or at known periods, and they can only be predicted probabilistically.

The retail banking context has elements of both contractual and non-contractual settings (called semi-contractual setting by Borle *et al*. (2008)), as we have noted in the previous paragraph, a scenario that has not been analysed in-depth previously (Singh and Jain, 2007): customers make random transactions (for example, credit card transactions) and have a contract for each banking product purchased. In accordance with the data available, we have assumed a bank as a multi-product or multi-service context because it offers different kind of products to customers (e.g., accounts, credit cards, debit cards, different types of insurances, loans, mortgages and deposits). We have at our disposal longitudinal data about the product portfolio owned by each customer and not about specific transactions performed by customers (for example using credit or debit cards). Therefore, for our research and due to the limitations of the data at our disposal, the retail-banking context is configured as a **contractual setting**. This choice is considered as the best way to proceed if we take into account our main premise, which is to develop an integrative model that bears in mind the whole portfolio of products (and not just a subset of them) that each customer has acquired for a period of time. Working with transactions of different kinds of products would be more difficult and unmanageable, because each service has different characteristics (e.g., the periodicity of transactions of a home insurance is a year and this periodicity is undetermined for credit or debit cards). Additionally, the bank observes the time at which a customer becomes inactive (attrition) because the company maintains a record of each customer by establishing contracts with them, therefore retention rate (or its opposite, churn rate) is a directly observable variable. In conclusion, although we have noted that some authors had assumed concrete bank products as non-contractual[6] or even semi-contractual, we had to choose the contractual setting due to the particularities of working with a portfolio of company products, not only one (or a subset of them). Thus, a contractual setting is the best way to tackle our problem.

3.1.2. Historical or predictive analysis?

Starting from the premise that the past dictates the future, there are two different ways to build a CLV model: (i) *historical* and (ii) *predictive* analysis (Jackson, 1989b). Kumar and George (2007) call this approach to classify studies: *the time period of calculation*. For Kumar and George (2007), it can be finite (which corresponds to historical models) or infinite (which corresponds to predictive models).

---

[6] Such is the case of Glady et al. (2009a), who only analysed checking accounts which were considered as non-contractual, because even if the general relationship is long and contractual, the price for the customer to stop using it is low and the product usage is at the customer's discretion. Other examples are: Haenlein et al. (2007) and Glady et al. (2009b).

The first group of models, **(i) *historical CLV models***, which are based on customer data available, examine only what happens in the past (Reinartz and Kumar, 2003; Venkatesan and Kumar, 2004). The second group of models, **(ii) *predictive CLV models***, as a result of historical perspective, they want to discover what will happen in the future under similar conditions (Gupta *et al*., 2004; Malthouse and Mulhern, 2008).

The models that try to calculate the long-term value of the financial contributions of a customer always include a retention rate, a time horizon of the study, or both. Since retention rates are generally less than one, some researchers state that the research time horizon should be infinite (Gupta *et al*., 2004). In theory, CLV models should estimate the value of customers across the entire customer-company relationship (Benoit and Van den Poel, 2009), although in practice using a finite time period of data from three to four years (Donkers *et al*., 2007; Rust *et al*., 2000), or even shorter time periods, seem to be sufficient to capture the possible changes in the environment (two or three years are suggested by Kumar (2008b) and one year is period of data available for Hwang *et al*. (2004)). However, the longer the span of period over which the data is collected the better is (Kumar, 2008b p. 81) and the goal should be to work with a data period that is broad enough to reflect the reality of the marketplace. Therefore, our challenge is how to **predict** the future cash flow of a customer based on their past behaviour and for this task we work with **2 years of data (24 months) from a Spanish retail bank**. In particular, the observed time period of data comprises from December 2010 to November 2012.

3.1.3. Deterministic or stochastic analysis?

Gupta *et al*. (2006) identify six types of models that researches have usually used to examine the CLV components (acquisition, retention and expansion or cross-selling). These models are: RFM models, probabilistic models, econometric models, persistence models (multivariate time series analysis), computer science models (data mining, machine learning and nonparametric statistics), and growth and diffusion models. In particular, the ***deterministic equations*** in which the terms are entered directly in the calculation of CLV are used in the first analysis (Dwyer, 1997; Berger and Nasr, 1998; Blattberg and Deighton, 1996). These models adopt simplified calculations that ignore heterogeneity of individual customer response probabilities (e.g., customers' retention and/or customers churn rates within a cohort), producing formulas that can be easily used by managers and solving a greater number of managerial problems, but in a way purely descriptive. The deterministic models include (i) *RFM models* and (ii) *growth and diffusion models* (for a review see Gupta *et al*., 2006):

(i) ***RFM models*** describe customer behaviour based on three variables of customer past buying behaviour or prior purchases, they are: recency (time since the last transaction), frequency (number of transactions during a time period of calculation) and monetary value (of transactions). The simplest models classify customers into groups based on each value of these three variables (e.g., Kahan, 1998; Marcus, 1998; Miglautsch, 2000). In the same vein, other studies use weights to each RFM variable to assign different levels of importance to these RFM variables (e.g., Hu and Jing, 2008; Liu *et al*., 2011). All these simplest models suffer several limitations and they have been criticised by some researchers mainly because they predict the behaviour only for the next period, RFM variables are considered imperfect indicators of actual behaviour, ignore the effect of other variables, such as marketing activities undertaken by the company and do not offer the monetary amount of customer value as a model output (Reinartz and Kumar, 2003; Venkatesan and Kumar, 2004) (for more details see section 4.1.1, where you can find a review about RFM variables). In spite of these limitation, measuring RFM variables is an important method for assessing CLV (Liu and Shih, 2005a, 2005b) and these limitations have been overcome incorporating RFM in CLV models through stochastic modelling approaches (e.g., Abe, 2009b; Borle *et al*., 2008; Fader *et al*., 2005b; Pfeifer and Carraway, 2000), because RFM models provide enough statistical rigor to serve as a basis of a CLV model (Fader *et al*., 2005b). Pfeifer and Carraway (2000) were one of the first researchers that solved these limitations mixing RFM models with CLV concept through Markov Chain Models (MCM). Later, other authors have used other methodologies to mix RFM measures with the CLV concept, as Pareto/NBD models (Fader *et al*., 2005b), and through a Hierarchical Bayesian approach (Abe, 2009b; Borle *et al*., 2008).

(ii) ***Growth and diffusion models***, such as the Bass model, using aggregated data to describe the number of customers who are likely to acquire by company in the future (Gupta *et al*., 2004) or the direct value (profitability) and indirect value (word of mouth) of lost customers by the company (Hogan *et al*., 2003), among other applications.

A ***stochastic process*** is used to characterise a sequence of random variables (stochastic) that evolve in terms of another variable, usually time. Each of the random variables of the process has its own probability distribution function, and among them, they may or may not correlate. Stochastic CLV models bring much more precision to CLV calculations by considering customer heterogeneity (e.g., in retention and/or in churn rate). In the framework of stochastic modelling related to CLV, we could find four types of methodologies used by researchers to model the drivers or components of CLV, i.e., acquisition, retention and margin expansion (e.g., cross-selling and up-selling) (for a review see Gupta *et al*., 2006). These methods are: (i) *probability models*, (ii) *econometric models*, (iii) *persistence models* (time series analysis), and (iv) *computer science models*:

(i) A ***probability model*** is a representation of reality in which the observed behaviour is modelled as a stochastic process governed by an unobserved or latent behaviour, which is different among individuals according to some probability distribution. This is used to describe-predict behaviour. One of the first models in this category explicitly used to estimate the variable P(Alive) as a component of CLV in a non-contractual setting was the Pareto/NBD (Schmittlein *et al*., 1987; Reinartz and Kumar, 2000, 2003). Also within this category are Markov Chains (Libai *et al*., 2002; Rust *et al*., 2004a) used to create models of buying behaviour. Recently some researchers have used another type of probability model to estimate CLV called Hierarchical Bayesian approach (Abe, 2009b; Borle *et al*., 2008).

(ii) ***Econometric models*** share the same philosophy as probability models. In particular, hazard models estimate customer retention similar to the Pareto/NBD, but hazard models are applied to another context in which the duration of the customer-company relationship can be measured (contractual or lost for good) (Van den Poel and Larivière, 2004). When trying to model the change, for example between suppliers (if data on competitors are available), again the Markov chains are set up as a model to consider that also could be framed within this group (e.g., Rust *et al*., 2000, 2004a).

(iii) When you have enough time series data, ***persistence models*** make possible the processing of such data (e.g., VAR models, unit roots, cointegration). In particular, the VAR methodology has been used in the context of the CLV to study the impact of advertising, discounts and product quality on CE (Yoo and Hanssens, 2005) and to examine differences in CLV between customers acquired through different marketing channels (Villanueva *et al*., 2008).

(iv) The application of ***computer science models*** (such as data mining, machine learning and statistical non-parametric models) for the calculation of CLV (e.g., Neslin *et al*., 2006) is configured as a prolific research stream, since they are able to deal with large amounts of data (variables) providing results with high predictive ability. As we have previously shown, marketing discipline has traditionally paid more attention to parametric statistics (based on a theory and easy to interpret). Therefore, computer science models should be explored and exploited in the future.

For the purpose of this research, **we mix probability models** (i.e., hierarchical Bayesian model) **to calculate individual CLV and computer science models** (i.e., data mining techniques, such as decision trees) **to segment the customer base.**

3.1.4. Source of data

Villanueva and Hanssens (2007) propose a comprehensive typology of CE models based on the data source for analysis, with the following categories: (i) *internal databases*, (ii) *surveys*, (iii) *company reports* (public information), (iv) *panel data*, and (v) *managerial judgments*. The type of data available for each company often depends on the type of relationship with customers as well as determines the unit of analysis, as we explain below.

Companies with a **(i) *database of customers*** normally are those whose relationships with their customers are governed by a contract, and they have data about individual customers (Venkatesan and Kumar, 2004; Verhoef and Donkers, 2001). Additionally, a firm in a non-contractual setting can record customer transactions in a database to perform posterior analysis (e.g., Kumar *et al*., 2006a). This kind of information allows the calculations to occur at the individual level and subsequently to get individual CLV's.

If no information is available from databases, a **(ii) *survey*** that collects customer perceptions is another important source of information for each individual customer (Rust *et al*., 2004a), which allows to gather information even on competitors and implement a CLV model through modelling techniques less complicated than if we had a customer base. This type of data is configured as an important information resource for small businesses, which often have less access to database technologies.

Other sources of information are *public reports*, *panel data* and *managerial judgements*. If the company pursues only the objective of assessment, it is enough to have access to **(iii) *public reports***, such as financial statements (Gupta *et al*., 2004; Gupta and Lehmann, 2003). These data collect and aggregate information, enabling analysis at this level (i.e., aggregated CLV). When customers switch brands frequently, it is interesting to have collected **(iv) *panel data*** with the effect of competition at an individual customer level (Yoo and Hanssens, 2005). Finally, **(v)** ***managerial judgments*** themselves are also configured as a possible source of aggregate information (Blattberg and Deighton, 1996; Ryals, 2005).

For the purpose of this research, we work with a **longitudinal database from a Spanish retail financial services company**. Also we are going to be able not only to calculate the present value of customers from past data, but also are going to be able to predict their future value.

3.1.5. Is effect of competition included?

Despite including the effect of competition in the calculation of CLV could enrich the results (e.g., Reinartz *et al*., 2005; Yoo and Hanssens, 2005), most models have not explicitly included this information, because it is difficult to obtain (e.g., Villanueva *et al*., 2008; Ryals, 2005).

The effect of competition has been measured through perceptions, collected by customer surveys and modelled by Markov processes to study, for example, brand switching (Rust *et al*., 2004a), or by data panel and time series analysis (Yoo and Hanssens, 2005).

For the purpose of this research, **we do not have at our disposal this type of data from the financial services company database**, because the company does not collect this information about competitive firms.

3.1.6. Level of aggregation in the data for the CLV calculation

About this topic, two approaches have been developed for the assessment of customers. A company may (i) *calculate the total value of its customer base from aggregate financial measures* at a global level or by customer segments (e.g., Berger and Nasr, 1998; Blattberg *et al*., 2001a; Gupta *et al*., 2004; Rust *et al*., 2004a), or (ii) *calculate the value of each individual consumer from individual historical data* (e.g., Kumar and Shah, 2009; Lewis, 2005; Venkatesan and Kumar, 2004; Verhoef and Donkers, 2001). Based on the level of aggregation in the data, the estimation objectives could be diverse (for recent reviews about this topic see Kumar *et al*. (2004), Kumar and George (2007) and Malthouse and Mulhern (2008)).

When companies perform **(i)** ***calculation of the average CLV from aggregate measures***, the most common application has been to determine how much to invest in acquiring new customers, as well as retain existing ones. Such investments should not exceed the CLV (Blattberg and Deighton, 1996). Another important application is the estimation of the value of the customer base as an intangible asset of the company, in particular by assessing competitors using public data such as annual reports and financial statements (Gupta and Lehmann, 2003). Finally, firms use aggregate CLV to calculate the market value of a company with which to base decisions on mergers and acquisitions (Gupta *et al*., 2004).

On the other hand, when companies perform **(ii)** ***calculation of individual CLV from individual measures***, the most frequent applications have been the calculation of the duration of profitable lifetime of customers (Reinartz and Kumar, 2000) to obtain optimal methods of resource allocation to optimise CLV (i.e., prioritise and select customers based on the variables that explain differences in the duration of profitable lifetime of customers (Reinartz and Kumar, 2003)); or allocate marketing resources to individual customer, choosing the best mix and frequency of marketing contacts to each customer (Drèze and Bonfrer, 2002; Venkatesan and Kumar, 2004).

Therefore, CLV can be managed (i) at an *individual* or (ii) at an *aggregate* level. In the first case, the marketing actions depend on the individual customer value and in the second case marketing decisions are assessed based on their impact on the whole (global or segments) of the customer base (CE). Empirical studies have shown that customer value is usually not constant (Mulhern, 1999). In some cases, following the Pareto principle, 20% of customers can generate over 80% of profits (Stahl *et al*., 2003). Moreover, researchers frequently find that the top 20% of customers generate between 150% and 300% of total profits; the middle between 60% and 70% of customers just about break-even; and the bottom between 10% and 20% of customers makes the firm lose between 50% and 200% of total profits (Kaplan and Narayanan, 2001; Lingle, 1995). If the company loses the top 20%, it loses the most valuable customers. This fact will have a negative impact on the business and managers should know precisely which customers should be targeted for acquisition or retention efforts. For this reason and trying to avoid the loss of the most profitable customers for the firm, we **calculate the CLV with individual or disaggregated data of a financial service retailer**. Once we have individual CLV's, we perform an *ex poste* segmentation. For this purpose, we are going to take into account the individual customer values (CLV_i), because as some authors suggest, calculating the value of the firm and then segment the customer base is a way to enrich this kind of models (Bruhn *et al*., 2006; Keiningham *et al*., 2006; Kumar *et al*., 2009) and it is indispensable for optimising marketing investments (Tirenni *et al*., 2007).

In summary, multidisciplinary approach is needed to complement the models developed to date, establishing a dialogue between marketing and finance (Bauer and Hammerschmidt, 2005; Wiesel *et al*., 2008), as well as dialogue between marketing and the discipline of computer science

(Gupta *et al*., 2006; Rust and Chung, 2006), to integrate their modelling with the marketing measures. That is the reason that justifies our choice to develop a customer valuation model within the framework of CLV, dealing with individual data, as a tool to predict the future value of existing customers and using for this task: probabilistic modelling tools and data mining techniques, as we have explained in previous sections.

3.2. CLV and CE in the banking context

Because of the interest that the banking context has for this research, from the previous classification (see section 3.1.) here we explain in depth some studies that have developed CLV models in such context. Additionally, we describe models developed for an insurance context, due to the similarities between both types of companies.

With the objective of developing a framework that provides insights into the *potential value of current customers*, **Verhoef and Donkers (2001)** use data of current customers from an insurance company (a multi-service provider) to develop a model to predict the *potential value* of a current customer, i.e., the profit or value delivered by a customer if this one behaves ideally. The data were collected by surveys (to ask customers about relationship duration, purchase level of insurances, claiming behaviour, age, education, household size, income, and home ownership) and from the company database (with information on the purchasing behaviour of customers at the company, and some other characteristics, such as age and relationship duration). Specifically, Verhoef and Donkers (2001) compare a choice based model using univariate and multivariate Probit, with a potential value model, based on a linear regression model. The results can then be used as input for customer segmentation to improve segment specific strategies. In particular:

In the situation without dependence across different services (i.e., the errors are independent across individuals), Verhoef and Donkers (2001) explain that a (univariate) Probit model for purchases of product j, $j = 1, \ldots, J$, by customer i is adequate. It is specified in the formula enclosed below, where for i = 1, ..., N and j = 1, …, J: y_{ij}^{*} is an unobserved variable; y_{ij} is the ownership of product or service j for customer i (1 = ownership, 0 = no ownership (survey)); X_i are socio-demographic variables (e.g., age, income) of customer i (from customer information file or external); Z_{ik} is the observed ownership of product or service k at company for customer i (customer information file) and ε_{ij} is the error term:

$$y_{ij}^{*} = \beta_j X_i + \sum_{k=1}^{J} \gamma_{jk} Z_{ik} + \varepsilon_{ij},$$

$$y_{ij} = 1 \quad if\ y_{ij}^{*} > 0$$

$$y_{ij} = 0 \quad if\ y_{ij}^{*} \leq 0$$

$$Prob(y_{ij} = 1) = Prob(\varepsilon_{ji} > -\beta_j X_i - \sum_{k=1}^{J} \gamma_{jk}\, Z_{ik})$$

Additionally, Verhoef and Donkers (2001) enclose the following formula for the potential value:

$$Potential\ value_i = \sum_{j=1}^{J} Prob(y_{ij} = 1) * Profit_j$$

In many cases, purchase decisions are made simultaneously, or at least, they are related. Multivariate Probit model allows for correlations between the errors terms in the Probit equations for each service. Therefore, in the situation with dependence across different services Verhoef and Donkers (2001) explain that a multivariate Probit model for purchases of product j, $j = 1, \ldots, J$, by customer i is adequate. These authors get the following equation to compute the potential value of customer i, where $Prob(customer\ i\ owns\ portfolio\ k)$ is the probability of customer i purchasing portfolio k and $Profit_k$ is the Profit margin of all services in portfolio k:

$$Potential\ value_i = \sum_{k=1}^{K} Prob(customer\ i\ owns\ portfolio\ k) * Profit_k$$

When the researcher is solely interested in a customer's potential value itself, and not in the services that determine this potential value, a simple regression model can be used to predict the potential value of a customer. Predictions of potential value can then be based on an Ordinary Least Squares estimate of the following regression model:

$$Potential\ value_i = \beta X_i + \sum_{k=1}^{J} \gamma_k\, Z_k + \varepsilon_i$$

When interest is limited to a segmentation of the customer base into a high potential and a low potential segment, a suitable model that can be used is the Probit model for segment membership: the ordered Probit model. The Probit model for membership of the high potential value segment is defined as follows, where y_{ij}^* is an unobserved variable; y_i = 1 indicates that individual *i* is in the high potential value segment, while y_i = 0 indicates otherwise:

$$y_i^* = \beta X_i + \sum_{k=1}^{J} \gamma_k Z_{ik} + \varepsilon_i,$$

$$y_i = 1 \quad if\; y_i^* > 0$$

$$y_i = 0 \quad if\; y_i^* \leq 0$$

In their empirical application, Verhoef and Donkers (2001) use a median split to segment the customer base into two equally sized parts. The estimation results for the Probit model for service purchases and the regression model for potential value are also used to segment the customer database into two segments of equal size, at least in the estimation sample. Therefore, potential value is used as input for customer segmentation, so companies can invest in the customers (segments) that are (potentially) valuable for the company, but also minimise their investments in non-valuable customers.

Verhoef and Donkers (2001) only consider current potential value, whereas other posterior authors (in other contexts, i.e., wireless communication industry) have used this same way to proceed to calculate CLV (e.g., **Hwang *et al.*, 2004; Kim *et al.*, 2006**). For these authors, CLV consists of *current value*, *potential value* and also they add *customer loyalty*. In particular, Hwang *et al*. (2004) suggest a new LTV model of individual customer considering these three components of CLV (the first summation refers to past profit contribution and the second one refers to expected future cash flow), where t_i is the service period index of customer *i*, $N_i$is the total service period of customer i, d is the interest rate, $E(i)$ is the expected service period of customer i, $\pi_p(t_i)$ is the past profit contribution of customer *i* at period t_i, $\pi_f(t_i)$ is the future profit contribution of customer *i* at period t_i and $B(t_i)$ is the potential benefit from customer *i* at period t_i:

$$LTV_i = \sum_{t_i=0}^{N_i} \pi_p(t_i)(1+d)^{N_i-t_i} + \sum_{t_i=N_i+1}^{N_i+E(i)+1} \frac{\pi_f(t_i)+B(t_i)}{(1+d)^{t_i-N_i}}$$

In the theoretical part of this research we have explained that relationships between companies and customers have three dimensions that should be considered in CLV formulas, called: (1) length, (2) depth and (3) breadth. **Verhoef (2004)**, in an empirical application of CLV using data from an insurance company (considered as contractual setting), imputes the underlying behaviours into the equation to estimate CLV in the following way, where $P(retention)_{i,t}$ is the probability of continuation of the relationship for customer *i* at time *t* (length of the relationship), $Product_{i,j,t}$ is the purchase of product or service *j* by customer *i* at time *t* (breadth), $Usage_{i,j,t}$ is the usage of product or service *j* by customer *i* at time *t* or amount of service purchased (depth), $Margin_{j,t}$ is the contribution margin for product or service *j* per usage or volume entity on time *t* and *d* is the discount rate:

$$CLV_{i,0} = \sum_{t=0}^{T} \frac{P(retention)_{i,t} * (Product_{i,j,t} * Usage_{i,j,t} * Margin_{j,t})}{(1+d)^t}$$

Once again using data from an insurance company, **Donkers *et al.* (2007)** study the capabilities of a range of models to predict CLV. The simplest models, called *relationship-level models*, can be constructed at the customer relationship level (aggregated across all services), they are: status quo model, profit regression, RFM for customer retention, Probit model, bagging approach, duration models and Tobit II model; and more complex models, called *service-level models*, are focused on the individual services, they are: univariate and multivariate choice modes and duration models. See Table 8, where Donkers *et al*. (2007) show the different competing formulas.

Table 8. Overview of rival models (*)

Model	Mathematical model
1. Relationship-level models	
1.1. Status quo	$\text{Profit}_{i,t+1}=\text{Profit}_{i,t}$
1.2. Profit regression	$\text{Profit}_{i,t+1}=\alpha_0+\alpha_1\text{Profit}_{i,t}$
1.3. Retention (segmented)	
Aggregate retention rate	$E_t\{\text{Profit}_{i,t+1}\}=P_t(ret_{t+1})\text{Profit}_{i,t}$
Segmented retention rate	
Recency of last purchase/ cancellation	
Frequency (number of insurances)	$E_t\{\text{Profit}_{i,t+1}\}=P_t(ret_{t+1}\mid \text{segment}_i)\text{Profit}_{i,t}$
Monetary value (high vs. low profit)	
RFM-index (high vs. low)	
1.4. Probit model	$E_t\{\text{Profit}_{i,t+1}\}=P(\text{ret}_{t+1}\mid x_i)\text{Profit}_{i,t}$
1.5. Bagging	$E_t\{\text{Profit}_{i,t+1}\}=P(\text{ret}_{t+1}\mid x_i)\text{Profit}_{i,t}$
1.6. Duration model	$E_t\{\text{Profit}_{i,t+1}\}=P(\text{ret}_{t+1}\mid x_i, \text{Duration}_{i,t})\text{Profit}_{i,t}$
1.7. Tobit II model	$E_t\{\text{Profit}_{i,t+1}\}=\Phi(x_i'\beta)(\alpha_0+\alpha_1\text{Profit}_{i,t}+E\{\varepsilon_{i,t+1}\mid \text{retention}\})$
2. Service-level models	
2.1 Independent choice models	$E_t\left\{\text{Profit}_{i,t+1}\right\}=\sum_{j=1}^{J}P\left(\text{purchase}_{i,t+1,j}\mid x_i\right)\text{Margin}_j$
2.2 Independent duration models	$E_t\left\{\text{Profit}_{i,t+1}\right\}=\sum_{j=1}^{J}P\left(\text{purchase}_{i,t+1,j}\mid x_i,\text{Duration}_{i,t}\right)\text{Margin}_j$
2.3 Multivariate choice models	$E_t\left\{\text{Profit}_{i,t+1}\right\}=\sum_{y\in Y}P_t\left\{\text{portfolio}_{i,t+1}=y\mid x_i\right\}*\text{Margin y}$
2.4 Multivariate duration models	$E_t\left\{\text{Profit}_{i,t+1}\right\}=\sum_{y\in Y}P_t\left\{\text{portfolio}_{i,t+1}=y\mid x_i,\text{Duration}_{i,t}\right\}*\text{Margin y}$

(*) Source: Donkers *et al*. (2007)

In line with the traditional CLV literature (e.g., Berger and Nasr, 1998), Donkers and his colleagues only include past behavioural data available from the customer database to predict customer behaviour, since RFM variables of a customer have proven to be powerful predictors of future behaviour (Rossi *et al*., 1996). Donkers *et al*. (2007) include dummy variables for ownership of each insurance type in the previous year, which represent purchase frequency, but also profit (monetary value), as the sum of the purchase dummy variables for each type of insurance times the insurance specific margins. For recency, these authors include two dummy variables: one dummy indicating whether the customer purchased a new service the last year (purchase recency), and other dummy indicating whether the customer cancelled a service last year (cancellation recency). Donkers *et al*. (2007) also include relationship age as predictor of customer behaviour. In the regression-type models they include two dummy variables for the first and second year of the relationship. The duration models include relationship length only through the hazard rate. Finally, loyalty program membership is included as a predictor in the models (Bolton *et al*., 2000). After estimating the models and computing the CLV predictions, the predictive performance of them is compared in three domains: (1) predicting the level of individual CLV, (2) predicting a ranking of customers based on CLV, and (3) predicting the value of the total customer base. Finally, the main conclusions of this research are: (1) simple models perform well, and (2) focusing only on customer retention is not enough, cross-buying (breadth dimension) needs to be accounted for, as we have proposed for this research.

Haenlein *et al*. (2007) also provide a contribution in the CLV area by presenting a customer value model focuses on the assessment of homogeneous segments of customers instead of individual customers. Haenlein *et al*. (2007) develop this model in cooperation with a leading German retail bank, considered as non-contractual setting. The model is based on four different groups of profitability drivers: age, demographics/lifestyle data, product ownership (type and intensity) and activity level. These profitability drivers have been defined in cooperation with the management of the retail bank that provides the data. These authors also measure contribution margin, defined as revenue resulting from interest payments and commission fees less liquidity cost, equity cost, risk cost and transaction cost covering the cost of holding cash of the bank, maintaining a certain credit risk-dependent equity ratio, accepting the risk of credit loss and carrying out customer-related transactions, respectively.

Haenlein *et al*. (2007) implement a combination of CART (classification and regression tree) and first-order Markov chain model. On one hand, CART analysis (with contribution margin as target variable and the aforementioned profitability drivers as predictor variables) is used to obtain homogeneous groups of customers. On the other hand, using these subgroups as discrete states, a transition matrix is estimated, which describes movements between groups in next periods. In a final step, this transition matrix is used to calculate CLV for each of the homogeneous subgroup of customers. In particular, Haenlein *et al*. (2007) determine the CLV for each customer group as the discounted sum of state dependent contribution margins, weighted with their corresponding transition probabilities. This calculation was carried out using backward induction.

The model can deal equally well with discrete one-time transactions as with continuous revenue streams. Furthermore, it is based on the analysis of homogeneous groups instead of individual customers and it is easy to understand and parsimonious in nature. Despite these benefits, the authors recognise several methodological limitations, for example they assume customer behaviour to follow a first-order Markov process, where the transition probabilities depend only on the behaviour during the last period, and their analysis is built on the assumption that the transition matrix will be stable and constant over time. This may be appropriate for medium-term forecasts, but might not be a sensible assumption for long-term forecasting. Moreover, we note that Haenlein *et al*. (2007) only measure retention through activity level variable, avoiding using the famous RFM framework, which could enrich this measure about behavioural loyalty.

Because of the fact that Haenlein *et al*. (2007) assume marketing budgets to be constant for all customers, they suggest for further research to combine their approach with the work on CE of Rust *et al*. (2004a), by relaxing this assumption (i.e., constant marketing budget) and estimating the effect of customer specific marketing activities on the transition probabilities between the different states of nature. In particular, it would be possible to determine the potential increase or decrease in individual or aggregated CLV resulting from these marketing activities.

Recently, **Benoit and Van den Poel (2009)** analyse CLV by means of quantile regression using data from a financial service company, but only for a particular service (i.e., data of insurance policies). This technique extends the well-known mean regression model to conditional quantiles of the response variable, such as the median, providing a more nuanced view of the relationship of the dependent variable (CLV) and the covariates or independent variables (past behavioural data and socio-demographics), since it allows the user to examine the relationship between a set of covariates and the different parts of the distribution of the response variable.

Benoit and Van den Poel (2009) compute CLV using a basic formula to calculate CLV for customer i at time t for a finite time horizon T (Berger and Nasr, 1998), where d is a pre-determined discount rate:

$$CLV_{i,t} = \sum_{\tau=0}^{T} \frac{Profit_{i,t+\tau}}{(1+d)^{\tau}}$$

In multiservice industries, J is the number of different services sold, $Serv_{ij,t}$ is a binary variable indicating whether customer i purchases service j at time t, $Usage_{ij,t}$ is the amount of that service purchased and $Margin_{ij,t}$ is the average profit margin for service j:

$$Profit_{i,t} = \sum_{\tau=0}^{J} Serv_{ij,t} * Usage_{ij,t} * Margin_{ij,t}$$

Benoit and Van den Poel (2009) show the benefits of using quantile regression for modelling CLV, because this technique provides insights into the effects of the covariates that are missed by traditional least-squares estimates (e.g., linear regression) and provides prediction intervals that give insights into the uncertainty about CLV predictions. Moreover, these authors propose a segmentation scheme based on the combination of the CLV predictions and the uncertainty associated with the predictions. They do not distinguish between high and low CLV customers because they prefer to determine the exact distinction between certain and uncertain predictions.

Finally, Benoit and Van den Poel (2009) explain several shortcomings and future research streams. One of them is to apply the proposed methodology in other settings, to develop finer market segmentations using the proposed segmentation scheme, or extend the proposed scheme with some other variables proposed in previous research. Additionally, these authors measure profits, retention (through R and F) and demographic variables, although they do not measure, for example, both the cost derived to serve customers and cross-selling to check their impact on customer value.

Glady *et al.* (2009a) develop another recent example of CLV model, whose purpose is not to provide a new way to model CLV, or a new classification technique. Instead of this, their purpose is to provide a framework for assessing churner classification techniques based on financial measure accuracy, i.e., use a profit-sensitive loss function for the selection of the best classification techniques with respect to the profits loss incurred by a misclassification, considered from a CLV perspective (detect the customers with a decreasing loyalty, who are defined as those with a decreasing future CLV). To achieve this goal, several classifiers for churn prediction are used and compared: a decision tree and a neural network with a baseline logistic regression model.

This model takes into account only data about current account transactions (number of invoices last month, amount invoiced last month, number of withdrawals, etc.). All transactions are aggregated at the customer level. Glady *et al*. (2009a) consider two different product usages, the total number of debit transactions and the total amount debited by month. Credit transactions, for simplification purposes, are not taken into account.

In order to get the value of a customer, Glady *et al*. (2009a) firstly define CLV as the present value of future cash flows yielded by the customer's product usage, without taking into account previously spent costs. The CLV is a function of all the transactions a customer will make, for the

q products the company is selling, but it does not take into account cross-individual (word of mouth) effects. Consequently, the CLV of the customer i, for the horizon h from period t is calculated as we indicate below, where $CF_{i,j,t+k}$ is the sum of the net cash flows, yielded by the transaction on product *j*, discounted at the rate *r*. Since Glady *et al.* (2009a) are focused on retention and not acquisition, all customers were acquired in the past and only marginal earnings are to be accounted, disregarding acquisition cost and any sunk or fixed costs (the marginal profit considered is nearly equal to the transaction price paid by the customer, since the marginal costs of the transactions are negligible). Hence, Glady *et al.* (2009a) denote the marginal profit by unit of product usage for product *j* as π_i, assumed fixed by product (for simplicity reasons), and they define the net cash flow $CF_{i,j,t+k}$ generated by a product *j* sold to a customer *i* during period *t* as a function of the product usage $x_{i,j,t+k}$:

$$CLV_{i,t} = \sum_{k=1}^{h}\sum_{j=1}^{q}\frac{1}{(1+r)^k}CF_{i,j,t+k} = \sum_{k=1}^{h}\sum_{j=1}^{q}\frac{1}{(1+r)^k}\pi_i x_{i,j,t+k}$$

Subsequently, to detect churning behaviour (a churner is defined as someone with a CLV decreasing over time), Glady *et al.* (2009a) estimate the slope of the customer life-cycle, giving an insight on future spending evolutions. Combining these two ideas, these authors predict churn on the basis of the slope of CLV in time, moving from a product-centric viewpoint to a customer-centric one.

Glady *et al.* (2009a) show that the cost-sensitive approach achieves very good results in terms of the defined profit measure, emphasizing that, besides achieving a good overall classification, it is important to correctly classify potentially profitable churners. Despite this important contribution, these authors identify different topics for further research, for example a more accurate prediction of the CLV. They also study only non-contractual product types, without taking into account either cross-product effects (cross-selling), or cross-individual effects (word of mouth).

Glady *et al.* (2009b) also develop another recent example of CLV model in a banking context as a non-contractual setting, considering only customers that have purchased (or sold) stocks, bonds, mutual funds, derivatives, etc., and not considering the automated pension plan transactions, stock exchange transactions, and insurance product. Glady *et al.* (2009b) propose a modified Pareto/NBD model (Pareto/Dependent model) in order to take advantage of Fader *et al.*'s (2005b) proposed future research streams and finally to estimate CLV. In particular these Fader *et al.*'s (2005b) proposal refers to not relying on the independence assumption between the number of transactions a customer makes and the average profit yielded by these transactions. Consequently, Glady and his colleagues estimate CLV of the customer i for the horizon h as is indicated below, where d is the discount rate, assumed to be constant (it is taken as the weighted average cost of capital disclosed in the 2004 financial statement of the Belgian financial service institution, 8.92% yearly, 0.7146% monthly); and $Cash\ Flow_{i,k}$ is the net cash flow (i.e., the total gains less the total costs) due to the activity of customer i during the time period k. The CLV of a customer is obviously changing over time. Nevertheless Glady *et al*. (2009b) do not introduce this time dependency in the notation, since in their empirical study the moment of prediction of the CLV is identical for all customers:

$$CLV_{i,h} = \sum_{k=1}^{h} \frac{Cash\ Flow_{i,k}}{(1+d)^k}$$

Based on the previously shown formula, the CLV of customer i, computed for a horizon of h periods, is estimated under the Pareto/Dependent model by the following formula:

$$\widehat{CLV}_{i,h} = \sum_{k=1}^{h} \frac{\hat{x}_{i,T_{i+k}} m_{i,T_{i+k}} - \hat{x}_{i,T_{i+k-1}} m_{i,T_{i+k-1}}}{(1+d)^k}$$

The authors have assumed that the number of transactions and the average profits per transaction of a customer i are related by the model (i.e., the idea is that the monetary value of a customer depends on the number of transactions he/she is making), being r_i a coefficient of dependence:

$$\log\left(\frac{m_{i,k}}{E[m_{i,k}]}\right) = r_i \log\left(\frac{x_{i,k}}{E[x_{i,k}]}\right) + \varepsilon_i$$

Glady *et al*. (2009b) model r_i as a function of explicative variables, for which they take the cohort (T_i), recency (t_i) and the probability of being an active customer ($\hat{p}_i$). Estimating this regression equation yields estimates for the parameters α_0, α_1, α_2, α_3, and hence also an estimate for the r_i:

$$r_i = \alpha_1 \hat{p}_i + \alpha_2 T_i + \alpha_3 t_i + \alpha_0$$

Then, for $k = T_i$ and following the equation that relates the number of transactions and the average profits per transaction:

$$\log\left(\frac{\widetilde{m}_i}{E[m_{i,k}]}\right) = \alpha_1 \hat{p}_i \log\left(\frac{x_i}{E[x_{i,k}]}\right) + \alpha_2 T_i \log(\frac{x_i}{E[x_{i,k}]}) + \alpha_3 t_i \log\left(\frac{x_i}{E[x_{i,k}]}\right) + \alpha_0 \log\left(\frac{x_i}{E[x_{i,k}]}\right) + \varepsilon_i$$

The average profits per transaction in period $[0, T_i + k]$ can then be estimated as well:

$$\log\left(\frac{m_{i,T_{i+k}}}{E[m_{i,k}]}\right) - \log(\frac{\widetilde{m}_i}{E[m_{i,k}]}) = r_i \log\left(\frac{x_{i,T_{i+k}}}{E[x_{i,k}]}\right) - r_i \log\left(\frac{x_i}{E[x_{i,T_i}]}\right)$$

This yields as prediction of $m_{i,T_{i+k}}$, with x_i the number of transactions in the past (frequency), T_i the cohort of customer i and $\hat{x}_{i,T_{i+k}}$ as prediction for $x_{i,T_{i+k}}$ (see more details to estimate $\hat{x}_{i,T_{i+k}}$ in Glady *et al.*, 2009b p. 2065):

$$m_{i,T_{i+k}} = \widetilde{m}_i (\frac{\hat{x}_{i,T_{i+k}}/E[x_{i,T_{i+k}}]}{x_i/E[x_{i,T_i}]})^{\hat{r}_i}$$

The profit of a transaction is computed, by a business rule, as a margin of 1% of the amount exchanged at the transaction. When computing the CLV, Glady *et al.* (2009b) operate with monthly time periods. In addition, these authors measure retention through RFM variables, and as we have noted previously, they consider average profits per transaction dependent of the number of transactions.

Glady *et al.* (2009b) show that the newly proposed method has better forecasting performance than the traditional Pareto/NBD model, and that it also outperforms a standard regression approach. The main drawback of this research is that only transactional data were considered. One possibility for further research is to include, for example, socio-demographic explanatory variables for the CLV prediction.

3.3. Concluding remarks

In this chapter we have described the main characteristics of the most important CLV models in banking and insurance contexts and also the future research streams that each study proposes. These studies strengthen our framework to model CLV in a similar or (in some cases) even the same context (i.e., a multi-service context) and also help us to justify the choice of variables for our model, which we explain in detail in Chapter 4.

Chapter 4. DRIVERS AND COMPONENTS OF CLV-CE

Persson and Ryals (2010) make an important distinction between components and drivers of CE and, by extension, of CLV. Firstly, they point out that the components of CLV and CE are **retention rate**, **cash flows** (or alternatively profits) the firm expects to receive from the customer in each future period and the **discount rate**. We add to these three components two more. On one hand, the **level of cross-buying** of each customer or in other words, the portfolio of banking products that each customer chooses and purchases. On the other hand, the **level of usage** of each banking product (following the suggestions of Bolton *et al*. (2004) and Verhoef (2004) about CUSAMS framework).

Therefore, for our attempt to model CLV, the value of a customer of a multiservice retailer depends on three core behaviours (Bolton *et al*., 2004; Verhoef *et al*., 2001). For these companies, customer retention by itself is not fully responsive to the goal of value creation and we consider:

(1) The duration of the provider-customer relationship (length of relationship).

(2) The number of different services bought from the same provider (breadth of the relationship).

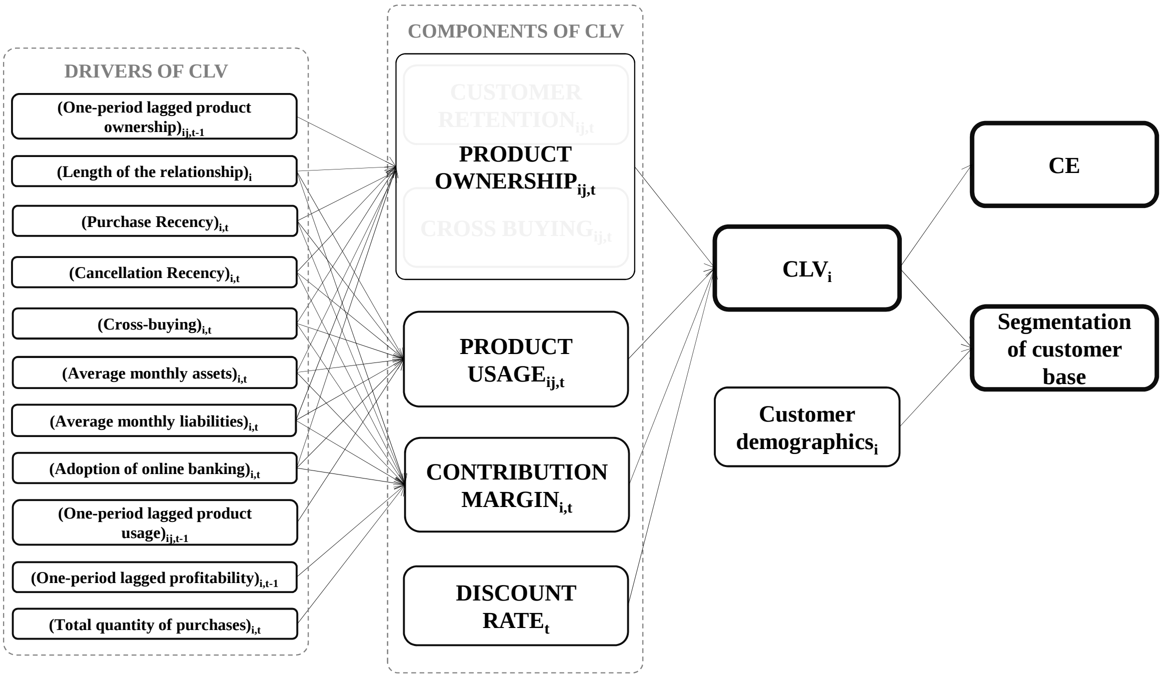
(3) The usage level of the consumed services (depth of the relationship).

Figure 3 shows a graphical representation of our proposed model, where length and breadth dimensions are represented by product ownership, depth dimension by product usage, also contribution margin and finally, the discount rate. To better understand this Figure, where retention is absent, it is essential to remark that some authors have noted that when one predicts the purchase or choice probability of each product (as is our case with product ownership), the retention probability does not have to be estimated separately (Donkers *et al*., 2003; Verhoef, 2004 p.23). Proceeding in this way we account for both customer retention and cross-buying.

Cross-buying has been associated with higher levels of customer retention, revenue generation and loyalty (Donkers *et al*., 2003). It has a significant positive influence on the relationship between the customer and the firm (Van den Poel and Larivière, 2004), because customers that acquire more products from the same firm experience increasing switching costs and are more likely to stay with the firm (Kamakura *et al*., 2003). Managers have increasingly acknowledged that retaining customer is not enough to be successful and many are seeking to enhance the value of their customers by expanding the range of products and services they buy from the firm (Blattberg and Deighton, 1996; Rust *et al*., 2000 p.46). Therefore, this link between retention and cross-selling justifies that some authors have noted that when one predicts the purchase or choice probability of each product (called in our model ***predictions of PRODUCT OWNERSHIP***$_{ij,t}$), the retention probability does not have to be estimated (Donkers *et al*., 2003; Verhoef, 2004 p.23). In short, when the purchase of a product by a customer is occurred, this concerns a contract renewal decision, thus this customer will be retained by the company.

Secondly, Persson and Ryals (2010) complement the CLV concept with its drivers. As we have noted previously, drivers of CLV can be customer perceptions and customer behaviours. In our research, we have focused on customer behaviours as drivers of CLV because there are several disadvantages related to using customer perceptions. Indeed, perceptions are ambiguous and backward-looking measures, and also there are measurement problems such as the common-method bias, response consistency and the discrepancy between what customers say and what customers do (Bolton *et al*., 2004; Chandon *et al*., 2005). In this chapter, we present and justify the set of variables that we use as drivers and components of CLV.

Figure 3. Drivers and components of CLV selected for this research (*)



(*) where *i* is the customer index, *t* is the time period index, *(t-1)* refers to the previous period of data, and *j* is the banking product index.

4.1. The first component of CLV: *PRODUCT OWNERSHIP*$_{ij,t}$

As we noted previously, some authors have suggested that when one predicts the purchase or choice probability of each product (in Figure 4 is called ***predictions of PRODUCT OWNERSHIP***$_{ij,t}$), as is our case using a hierarchical Bayesian model, the retention probability does not have to be estimated (Donkers *et al*., 2003; Verhoef, 2004 p.23). Proceeding in this way, we account for both customer retention and cross-buying. This is because when the purchase of a certain service by a certain customer occurs, this implies a contract renewal decision. Thus, this customer will be retained by the company. In this section we review why these two components are essential in order to calculate CLV.

4.1.1. Customer retention

The first component of CLV is ***customer retention***. Jacoby and Kyner (1973 p.2) define customer loyalty (or in other words, customer retention) by using a set of six necessary and collectively sufficient conditions: customer loyalty is *"(1) the biased (i.e., non-random), (2) behavioural response (i.e., purchase), (3) expressed over time, (4) by some decision-making unit, (5) with respect to one or more alternative brands out of a set of such brands, and (6) is a function of psychological (decision-making, evaluative) processes*". The authors state that the evaluation process (the sixth condition) makes an individual to develop a commitment towards a brand. Later, Oliver (1999 p.34) defines customer loyalty as a "*deeply-held predisposition to re-patronise a preferred brand or service consistently in the future, thereby causing repetitive same brand purchasing, despite situational influences and marketing efforts having the potential to cause switching behaviour*". When a customer is loyal (or retained), he or she continues to buy the same brand, tends to buy more and is willing to recommend the brand to others (Hepworth and Mateus, 1994).

Since early research about customer loyalty, at least three perspectives have been proposed in defining and operationalizing the concept: (i) the *attitudinal approach*, (ii) the *behavioural approach* and (iii) the *composite approach* (Jacoby and Chestnut, 1978). (i) The *attitudinal approach* goes beyond overt behaviour and expresses loyalty in terms of consumer's strength of affection towards a brand (Backman and Crompton, 1991; Bennett and Rundle, 2002). Attitudes have been related to behaviours, although it is important to differentiate that one may hold a favourable attitude toward a brand but not purchase it over multiple occasions because of comparable or greater preference toward other brands (Dick and Basu, 1994). Similarly, Sharp *et* *al*. (2002) suggest that attitude is not relevant to determine brand loyalty, because it is at odds with basic epistemological principles, it is not a stable measure (i.e., attitude adds another factor of uncertainty, rather than contributing to explanatory and predictive ability), and there is a certain contradiction in the existing causal explanations that relates attitude and brand loyalty in a variety of fields, including marketing. (ii) The *behavioural approach* defines loyalty strictly from a behavioural perspective. The major assumption here is that (repeat) purchasing will capture the loyalty of a consumer towards the brand of interest (Hughes, 1996a; Farley, 1964). Finally, (iii) the *composite approach* defines loyalty as consisting of repeated purchases prompted by a strong internal positive predisposition toward them (Day, 1969; Dick and Basu, 1994; Pritchard and Howard, 1997; Schijns and Schröder, 1996). This approach integrates both behavioural and attitudinal dimensions.

The concept of customer loyalty has always been at the forefront of firms that want to retain customers (Buckinx and Van den Poel, 2005; Grewal *et al*., 2004). To support this important stream, the attitudinal approach to measure customer loyalty is conceptually rich, but it is rather difficult for researchers to collect large-scale attitudinal data (Zhang *et al*., 2010). Consequently, in empirical research, attitudinal loyalty is not as widely used as behavioural loyalty (Uncles *et al*., 2003) because behavioural loyalty measures are more clearly observed than attitudinal measures. In addition, to measure loyalty it is more suitable to work with data from company databases, as is our case with the data available from a Spanish retail financial services company, because consumption records in such databases only imply customer behavioural loyalty (Wong and Chung, 2008). Therefore, in this research we are going to take into account the second approach to measure customer loyalty: *behavioural loyalty*.

To identify loyal customers from a behavioural perspective, firms have typically analysed customer behaviours with respect to the following questions (Kumar *et al*., 2006a):

a) For how long has the customer been active? (Colgate and Lang, 2001; Reichheld, 1996)

b) How regularly does the customer buy? (Farley, 1964; Massey *et al*., 1970)

c) What is the RFM score of my customer? (Hughes, 1996a; Kahan, 1998)

For the objectives of this research we have selected RFM variables and length of the relationship to measure behavioural loyalty in a complete way (Chang and Tsay, 2004; Kumar *et al*., 2006a; Li *et al*., 2011). As some authors have noted, customer retention is one of the key drivers of CLV and firm profitability (Gupta and Zeithaml, 2006). Therefore, through these measures customer retention is included in the model.

4.1.1.1. The use of RFM variables

RFM is a popular set of three variables that are defined as we explain below:

(1) Recency (R) or time since the last transaction or purchase. In a clearer way, this variable means the difference between the (period of) time that customer purchased the last time and the (period of) time of analysis. Some authors differentiate between purchase recency, as a binary variable indicating whether the customer purchased a new service the last period, and cancellation recency, as a binary variable indicating whether the customer cancelled a service last period (Donkers *et al*., 2007).

(2) Frequency (F) or number of transactions or purchases during a time period of calculation.

(3) Monetary value (M) of transactions or purchases during a time period of calculation or total monetary sales (average by month, year, or since the beginning of the relationship).

Literature has accumulated so many uses of this set of variables and there is overwhelming evidence both from academically reviewed studies as well as from practitioners' experience that the RFM variables are an important set of predictors for modelling repeat purchasing (Baesens *et al*., 2002). In particular, RFM has been one of the most widely used methods to identify best customers for the past thirty years, especially in direct marketing (Hughes, 1996a). The RFM triad was developed to target marketing programs at specific customers with the objective to improve response rates, because past purchase behaviour of customers is a better predictor of their future purchase behaviour than, for example, demographics (typically used to get profiles of customers before using RFM variables). Demographics variables explain what people are (e.g., age, sex, and income), RFM measures explain what people do (e.g., when they buy, how often they buy, and how much they buy), which it is what we are trying to predict.

To justify the use of RFM measures about past buying behaviour, Hughes (1996a), and more recently Liang (2010), explain that there is an a priori reasoning and empirical evidence about customers who recently purchased from a marketer (recency), those who purchase many times from a marketer (frequency) and those who spend more money with a marketer (monetary value). These customers typically represent the most valuable customers for the firm and therefore they are the best prospects for new offerings. Despite the importance of the previously mentioned idea, the real power of this technique comes from combining them into a three digit called: 'RFM code'.

The simplest RFM models use these three variables as inputs for scoring models to prioritise and select individual customers (e.g., Hughes, 1996a; Kahan, 1998; Marcus, 1998). To apply this kind of models, three pieces of information are necessary in every customer record: the RFM variables described previously. The simple models sort the database by each value of RFM variables, coding the top 20% (in terms of R firstly, in terms of F secondly and finally in terms of M) as 5, and the less quintiles (again in terms of R, F and M) as 4, 3, 2 and 1. Everyone in the database has his/her own score related to R, F and M and all customers are presented by 555, 554, 553… 111. These scores create 125 (5×5×5) RFM cells. Moreover, the best customer segment is 555, while the worst customer segment is 111. With this reasoning you could obtain an easy way to segment the customer base, for example to guide different direct communication campaigns to different groups of customers. More recently, some authors propose Weighted RFM (WRFM) instead of only RFM (e.g., Hu and Jing, 2008; Liu *et al*., 2011; Miglautsch, 2000). They allocate different weights to R, F and M depending on characteristics of the industry. To determine the relative importance (weights) of the RFM variables, the Analytic Hierarchy Process (AHP) method is used, although these weights depend on the expert's experience and therefore, are subjective (Liu *et al*., 2011).

The popularity of the simplest RFM approach is not surprising, given the limited information that is needed to score customers and assign them into groups and the simplicity of the technique. However, there are some serious drawbacks of this approach (Gupta *et al*., 2006; Fader *et al*., 2005b; Keiningham *et al*., 2006; Kumar *et al*., 2008b; Ryals, 2002). Particularly:

(1) RFM variables are imperfect indicators of their true underlying behaviours (Fader *et al*., 2005b; Gupta *et al*., 2006). Some authors, such as Fader *et al*. (2005b) draw RFM variables from a true distribution, but this aspect is completely ignored in the simplest RFM models.

(2) RFM assumes that how recently, how frequently, and how much a customer spends are the only three variables that determine the value of a customer. The real world shows us that there are numerous other alternative and/or supplementary factors (such as, objective quality, service quality, customer satisfaction, customer retention or customer loyalty),

that determine 'best' customers. These supplementary factors should be taken into consideration when identifying customers for acquisition or retention efforts (Keiningham *et al*., 2006), or in other words, help to predict the customer's future purchase behaviour and the customer value to the firm (Kumar *et al*., 2008b).

(3) RFM analysis is focused on revenue rather than cost and therefore does not capture the real profitability of a customer relationship (Ryals, 2002).

(4) Since it focuses solely on past behaviour, RFM analysis also suffers from the key drawback of all historical data, i.e., it fails to consider future potential or developmental growth (Keiningham *et al*., 2006) or in other words, it may not be reliable as a guide to the future.

(5) RFM is primarily a segmentation scheme, assigning customers to a group rather than calculating an individual score for each customer (Keiningham *et al*., 2006).

(6) Related to CLV, RFM methods are scoring models and do not explicitly provide an amount of money for customer value (Gupta *et al*., 2006).

Some researchers attempt to predict customers' future behaviour, i.e., response, with RFM measures using some type of regression method. For example, Berger and Magliozzi (1992) use regression analysis to determine the relative power of these three factors (as well as their possible interactions). Since the response is usually binary (respond/do not respond to a mailing, that is the same to buy/not buy), logistic regression is often used (e.g., Glady *et al*, 2009a; Lewis, 2006; McCarty and Hastak, 2007; Suh *et al*., 1999; Yang, 2004). The procedure starts with a test mailing. After receiving the results of these test mailings, logistic regression can be used to analyse the response variable as a function of several independent variables (they are not restricted to RFM variables, e.g., number of times purchased) and to provide an equation that can calculate the response probability for the entire customer database. The predicted variable is the response probability, which varies from zero to one, therefore the model can provide a probability of response for everyone in the database, given the estimated parameters for a set of predictors variables. However, this method does not provide discrete groups of people.

Despite the previously mentioned regression approach seems to predict quite well in practice, it is somewhat unsatisfactory for several reasons (Colombo and Jiang, 1999; Fader *et al*., 2005b; Kumar *et al*., 2008b). In particular:

(1) Models that are designed to predict well (i.e., regression models) often can result in a poor understanding and inability to distinguish correlates from drivers of the process leading to response. Despite the fact that theories of consumer behaviour are sometimes used to suggest which independent variables to include in a regression model, there is certainly a lack of theory in building regression models in the context of RFM approach (Colombo and Jiang, 1999).

(2) Regression models may be thought of as smoothing techniques that attempt to describe well the relationship between the predictors and the response but tend to treat heterogeneity as noise (Colombo and Jiang, 1999).

(3) Scoring models (based on RFM variables) are focused on past behaviour and they predict future behaviour only for the next period. To estimate CLV we need to estimate customer's purchase behaviour not only for period 2, but also for periods 3, 4, 5, and so on. It is not clear how a regression type model can be used to forecast the dynamics of buyer behaviour well into the future and then tie it all back into a present value for each customer (Fader *et al*., 2005b). Despite this disadvantage, Donkers *et al*. (2007) describe how to combine Markov Chain models with regression and choice models to predict CLV for an infinite time period.

(4) Two periods of purchasing behaviour are required: period 1 to define RFM variables and period 2 to arrive at values of the dependent variable(s). It would be nice to be able to leverage all the available data for model calibration purposes without using any of it to create a dependent variable for a regression-type analysis (Fader *et al*., 2005b).

When many other variables are available (in addition to RFM variables) and interactions are suspected, tree-based regression methods (e.g., AID, CHAID, CART) have been found to be useful in order to segment the customer base (e.g., McCarty and Hastak, 2007; Paauwe *et al*., 2007; Suh *et al*., 1999; Yang, 2004). In particular, Automatic Interaction Detection or AID and its relative, Chi-Square Automatic Interaction Detection or CHAID, are the most popular of a general class of techniques known as binary segmentation or 'tree' analyses in which segments are formed which exhibit the widest variation on the dependent variable. These methods are similar to the simplest RFM models because they create groupings (nodes) of database members. The main difference is that these groupings are not created a priori, such as is the case with RFM. Rather, the file is split according to a statistical algorithm after a test mailing is conducted. After the results of the test mailings are received, the procedure starts with a node that includes everyone in the database. The procedure then searches for the independent variable (e.g., number of times purchased) that best discriminates among the file members with respect to a dichotomous variable (i.e., purchased/did not purchase on current mailing). It splits the original node on this independent variable into as many subgroups as are significantly different (or discriminate each of them) with respect to the dichotomous variable. Through successive iterations into segments, the process ultimately arrives at 'terminal nodes' (because no other splits are significant and the terminal nodes are those that cannot be split any further), which represent segments with differing average response rates.

Finally, neural networks (NNs) have also been used to discover relationships between response and behavioural, demographic and other predictors (e.g., Baesens *et al*., 2002; Chan, 2005; Hsieh, 2004; Zahavi and Levin, 1997). NNs are mathematical representations inspired by the functioning of the human brain. A NN is typically composed of an input layer, one or more hidden layers and an output layer, each consisting of several neurons (layer units). Each neuron processes its inputs with the help of a propagation algorithm and generates one output value, which is transmitted to the neurons in the subsequent layer. The NN naturally produces a score per data input, which allows the data inputs to be ranked accordingly for scoring purposes (e.g., customer scoring). For decision purposes, the posterior probability estimates produced by the NN are used to classify the data inputs into appropriate (predefined) classes. In this context, from a predictive performance perspective, Bayesian NNs were found to be statistically superior when compared to logistic regression, linear and quadratic discriminant analysis classifiers (Baesens *et al*., 2002). However, NNs have associated also a number of disadvantages. For instance, networks are difficult to interpret (on the contrary, in regression models or hierarchical Bayesian models we can interpret the coefficients in relation to the problem), and convergence to a solution in NNs can be slow and depends on the initial conditions of the network (Warner and Misra, 1996 p. 292).

In the context of customer valuation, some researchers have compared the simplest RFM models with CLV models and have found CLV models to be superior (Reinartz and Kumar, 2003; Venkatesan and Kumar, 2004). Other researchers have settled the inherent limitations of the simplest RFM models mixing them with CLV concept (Fader *et al*., 2005b; Fader *et al*., 2007; Glady *et al*., 2009b; Kumar *et al*., 2006a; Pfeifer and Carraway, 2000). As we have explained previously, the reason that justifies this interest is that RFM might be good predictors of future purchase behaviour of customers and, at the same time, RFM statistics are sufficient to build a stochastic CLV model. Therefore, RFM variables can be used to build a more completed CLV model in a banking context than other CLV models developed to date (for more details see section 2.2 about CLV concept). As an illustration of this usefulness, Malthouse and Blattberg (2002) found that frequency of purchase is a good predictor of CLV.

Some authors have recently implemented a Bayesian approach that can naturally incorporate past behavioural outcomes (for example, RFM measures) into future expectations (Rossi and Allenby, 2003). Bayesian methods can incorporate such prior information in the structure of the model easily through the priors of the distributions of the drivers of CLV (Abe, 2009b; Borle *et al*., 2008). Furthermore, this approach can be used in any context. Therefore, we use such an approach to get CLV, which implies leveraging the extra information available to the firm in observing customer lifetimes.

4.1.1.2. Length of the relationship

We consider ***length of the relationship*** as another variable to measure customer loyalty. This concept refers to the duration of the relationship between customer and company or to customer retention (Bolton *et al*., 2004). Relationship length indicates a level of customer inertia that would be associated with greater loyalty (Bolton *et al*., 2004; Colgate and Lang, 2001; Reinartz and Kumar, 2000). Some of the researchers who propose the idea of customer relation length were Reinartz and Kumar (2000). They examine its influence on customer loyalty and profitability and suggest increasing relationship length to improve customer loyalty. Verhoef *et al*. (2002) also found a positive effect between the age of the relationship customer-firm and the number of services purchased in an insurance company (which is another indicator of an increasing profitability caused by increasing the length of the relationship between customers and firm).

Such is the importance of retention to the profitability of companies, reflected in a positive link between loyalty and firm profitability (Anderson *et al*., 1994; Reichheld, 1996; Silvestro and Cross, 2000), that many of them have implemented different types of loyalty programs to get relationships in the long term with customers, in order to strengthen the use of the product/service and to retain existing customers. Research about CRM has highlighted a great number of benefits that are associated with customers who maintain long-term relationships with companies (e.g., Reichheld and Sasser, 1990; Reichheld, 1996; Reichheld and Teal, 1996). In particular, the length of customer relationships influences the profitability of the firm, because the longer a customer stays the more he/she spends with the company (Reichheld, 1996). Moreover, Reichheld and Teal

(1996) attributed the increase in profits from loyal customers to the price premium paid by them, the added profits from sales through referrals, profits from cost savings obtained by serving an old customer, and revenue growth from a loyal customer due to the increase in sales to that customer. These benefits translate into the following five propositions (Van den Poel and Larivière, 2004):

(1) Successful customer retention lowers the need for seeking new and potentially risky customers and allows organizations to focus more accurately on the needs of their existing customers by building relationships (Dawes and Swailes, 1999; Engel *et al*., 1995).

(2) Long-term customers buy more (Paulin *et al*., 1998; Ganesh *et al*., 2000) and, if they are satisfied, may provide new referrals through positive word-of-mouth for the company (Ganesh *et al*., 2000; Colgate *et al*., 1996). In this line, some authors have also evidenced that customers that were referred to the company by friends, colleagues or family have higher CLV's (Kumar *et al*., 2010).

(3) Long-term customers become less costly to serve due to the great knowledge of the bank about these existing customers. This fact implies a decreasing in servicing costs (Ganesh *et al*., 2000; Paulin *et al*., 1998).

(4) They tend to be less sensitive to competitive marketing activities (Colgate *et al*., 1996).

(5) Finally, losing customers not only leads to opportunity costs because of reduced sales, but also to an increased need for attracting new customers (Athanassopoulos, 2000), which is five to six times more expensive than customer retention (Bhattacharya, 1998; Colgate and Danaher, 2000; Rasmusson, 1999).

Despite the fact that the previous proposals may seem very intuitive, some authors have shown that not all loyal customers are profitable (Reinartz and Kumar, 2002, 2003; Storbacka, 1997). The link between loyalty and profitability has been questioned in some industries (Reinartz and Kumar, 2002), mainly for two reasons: (1) a relatively large percentage of long-term customers are only marginally profitable, and (2) a relatively large percentage of short-term customers are highly profitable. However, Reinartz and Kumar's (2002) findings from four industries (high technology, postal service, retail food and direct brokerage) still indicate that a larger proportion of the long-term customers than of the short-term customers exhibit high profitability. Additionally, a larger proportion of the high-profitability customers than of the low-profitability customers are long-term customers. Thus, the theory of an overall positive connection between customer loyalty and profitability cannot be totally rejected.

In accordance with Anderson and Mittal (2000), customer relationship profitability is achieved by 'high quality' customers, who have low maintenance costs and high revenue. In the particular context of retail banking, Storbacka (1994) describes relationship costs and relationship revenue. The first one, relationship costs, comprise direct variable costs (such as transaction related costs and costs related to specific services), and overhead costs (that may or may not be attributable to particular relationships). The second one, relationship revenue, is split into volume-based revenue (that is derived from interest margins) and fee-based revenue. Since a large part of the revenues of a bank are received from interest margins, the customers' volume of business has a major impact on profitability. If relationship costs are minimised and relationship revenue is maximised over time, long-term customers should generate greater profitability than short-term customers. Finally, and making a bridge between loyalty and CLV, Reinartz and Kumar (2000) found that relationship length has a small correlation with future CLV in some contexts. Therefore, we are going to check if length of the relationship is a good (or bad) driver of CLV through its components.

4.1.2. Cross-buying

Another component of CLV is ***cross-buying*** (from the demand side) or ***cross-selling*** (from the supply side). Cross-selling aims to achieve the acquisition of a greater number of products from multiple categories by current customers of the firm (Gupta and Zeithaml, 2006). Whereas cross-selling is a firm-side action employed to broaden customer relationships, its counterpart on the demand side, cross-buying, refers to a customer's propensity to make cross-category purchases (Reinartz and Venkatesan, 2008). Cross-selling requires companies to offer the right products/services to individuals (Paas and Molenaar, 2005). It involves decisions such as assessing what products to offer, to whom, and when, to achieve such cross-selling (Kamakura *et al*., 1991; Kamakura *et al*., 2003; Knott *et al*., 2002). Customers who maintain long-term relationships with the company buy more (Paulin *et al*., 1998; Ganesh *et al*., 2000), and selling additional products to existing customers is much easier than attracting new customers (Felvey, 1982). This fact guides firms to optimise their assortments by extracting association rules among best-selling items. This task can be accomplished by means of the so-called *market basket analysis* (e.g., Brijs *et al*., 2004). It is a data mining technique which allows discovering which products are most likely purchased together. It is based upon the theory that if you buy a certain group of items, you are more (or less) likely to buy another group of items.

The importance of cross-selling also guides firms to predict acquisition sequences of products, for example through *next-product-to-buy (NPTB) models* (e.g., Prinzie and Van den Poel, 2007) or *acquisition pattern analysis* (e.g., Paas and Molenaar, 2005; Paas *et al*., 2007). The use of the first group of models (*NPTB models*) reduces the waste of poorly targeted cross-selling activities by predicting (using logistic regression, multinomial regression, discriminant analysis, or neural nets) the product (or products) each customer would be most likely to buy next, given what we know so far about the customer. These products can then be targeted to that customer. The second group of models (*acquisition pattern analysis*) investigates not only the order in which households acquire products, but also provides insight into timing of product acquisitions (for example, using latent class Markov models).

Cross-buying has been associated with higher levels of customer retention, revenue generation and loyalty (it depends on how behavioural loyalty is measured (e.g., share of wallet)). For example, Donkers *et al*. (2003 p. 21) found evidence that customers are more likely to stay when they purchase more products, in particular, more insurance types. When examining the relationship between cross-buying and behavioural loyalty, Reinartz *et al*. (2008) found that cross-buying is a consequence and not an antecedent of behavioural loyalty, because behavioural loyalty drives both the number of categories from which a person buys and the level of spending dispersion across those categories. Therefore, managers have increasingly acknowledged that retaining customer is not enough to be successful and many are seeking to enhance the value of their customers by expanding the range of products and services they buy from the firm (Blattberg and Deighton, 1996; Rust *et al*., 2000 p.46). Cross-selling has a significant positive influence on the relationship between the customer and the firm (Van den Poel and Larivière, 2004), because when customers acquire more products from the same firm, they experience increasing switching costs and, as we have noted previously, are more likely to stay with the firm (Kamakura *et al*., 2003). This clear link between retention and cross-selling justifies that some authors have noted that when one predicts the purchase or choice probability of each product (called in Figure 4 ***predictions of PRODUCT OWNERSHIP***$_{ij,t}$), the retention probability does not have to be estimated (Donkers *et al*., 2003; Verhoef, 2004 p.23). Proceeding in this way, as is the case for this study, we account for both customer retention and cross-buying (see Figure 3).

Customers with more products/services generate higher assets (Knott *et al*., 2002; Winer, 2001). Cross-selling increases the total value of the customer along his or her relationship with the firm (Kamakura *et al*., 2003). In particular, Jackson (1989a, 1989b, 1989c) shows how by offering six products (cross-selling and upgrades) to existing insurance policy owners during a 12 months period, an insurance firm increases the baseline CLV of an average customer by 40%. Even though a deterministic model is used, which calculates customer's value in an aggregate and very simple way, this is a good example that justifies the interest to study cross-selling.

Because of the increase in competition in the current markets, many firms are suffering a sales decrease, because (among other reasons) it is difficult for them to obtain new customers. Instead, selling products to existing customers through cross-selling strategies can increase company assets (Reichheld, 1996; Reichheld and Sasser, 1990). Cross-selling strategies have been emphasised during the last years because they help firms to maintain profits by selling a larger volume of products/services to existing customers (Prinzie and Van den Poel, 2007). In addition to the intensified competition and the critical situation that it generates for firms, the infrequency of purchasing financial services, sometimes due to their long lifetimes, magnifies the importance of cross-selling in the context of retail banking.

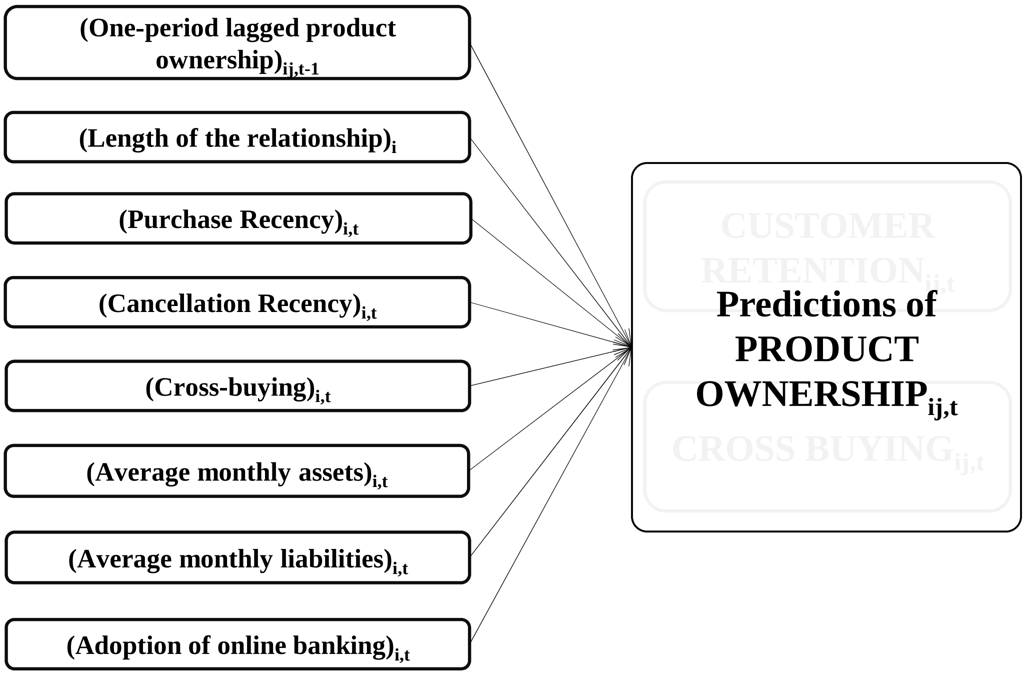
Moreover, the importance of including cross-selling in the research about the assessment of financial services customers is confirmed by authors as Kamarura *et al*. (2003 p. 47) and Prinzie and Van den Poel (2008 p. 714), who note that: "*cross-selling is effective for customer retention by increasing switching costs and enhancing customer loyalty, thus directly contributing to customer profitability and lifetime value (CLV)*". This idea is strengthened by the research of Villanueva and Hanssens (2007) and Gupta *et al*. (2006), who analyse the components and drivers of CLV, including as the most important ones: retention of existing customers and add-on selling (which is composed of cross-selling, up-selling, and higher quantity). In sum, cross-selling may be an important opportunity to improve CLV for some multiple service providers (Verhoef and Donkers, 2001).

4.1.3. Drivers of PRODUCT OWNERSHIP$_{ij,t}$

Therefore, we model the product choice *j* from a customer *i* in period *t* (called in Figure 4 ***predictions of PRODUCT OWNERSHIP***$_{ij,t}$) using a Bernoulli distribution (Ntzoufras, 2009 p. 257,263). This is because the response variable for all products in the database is binary, indicating with a 1 if a customer owns a product or 0 otherwise (for more details see Chapter 5).

For this task we have selected several drivers or antecedents of this product ownership (and by extension they are also drivers of CLV). In particular, there are: previous product ownership (one-period lagged variable), length of the relationship, recency variables (measured by purchase recency and cancellation recency), cross-buying variable, intensity of product ownership variables (measured by average monthly assets and average monthly liabilities) and adoption of online banking.

Figure 4. Drivers of product ownership (*)



(*) where *i* is the customer index, *t* is the time period index, (*t-1)* refers to the previous period of data, and *j* is the banking product index.

The antecedents that we adopt to predict product ownership are based on findings from previous research about this same topic (that is, product choice as a component of CLV). The first drivers that we include as predictors are: a one-period lagged variable that measures the ownership of each service (called here ***(one-period lagged product ownership)***$_{ij,t-1}$), the length of the relationship between customer and company (called here ***(length of the relationship)***$_i$, and recency variables (called as ***(purchase recency)***$_i$ and ***(cancellation recency)***$_i$). All of them have been previously used by other authors as predictors of product choice (Donkers *et al.*, 2007). In particular, about the inclusion of lagged values, they facilitate prediction and interpretation of causality. Therefore, we have decided to include this one-period lagged value as one of the antecedents in our analysis (Villas-Boas and Winer, 1999). Cross-buying (called ***(cross-buying)$_{i,t}$***) is also included as a predictor of product choice (Verhoef *et al*., 2001), mainly because creating value by cross-selling additional services is an important aspect of customer relationship management. Retaining customers is not enough to be successful and many authors and practitioners are seeking to enhance the value of the customers by expanding the range of products and services they buy from the firm (Blattberg and Deighton, 1996; Rust *et al*., 2000). Following the suggestions of Prinzie and Van den Poel (2006), we have also included intensity of product ownership or balances (called ***(average monthly assets)$_{i,t}$*** and ***(average monthly liabilities)$_{i,t}$***) as drivers of CLV. Prinzie and Van den Poel (2006) explain that past and current purchase behaviours are reflected by this (current) intensity of product ownership, and therefore this information is a good predictor of product choice and also of CLV (Haenlein *et al*., 2007; Reinartz *et al*., 2008). For a deeper understanding of the inclusion of these variables as predictors we also refer here to Haenlein *et al*. (2007). These authors define two conditions under which the customer is considered as active (versus inactive) in a banking context. These conditions are also applicable in the context of our study and we have adapted them according to the specific characteristics of the collaborating retail bank. In particular, these conditions are:

> *Condition 1*: All customers owning either a savings product, a home financing product, a loan or an insurance product are defined as being active due to the regular revenue streams (savings, interest payments, insurance fees) resulting from any of these products. To check this first condition (that is, whether or not a customer is active), these both pieces of information are used: type of product ownership and intensity of product ownership (measured by average monthly assets and average monthly liabilities). In particular, Prinzie and Van den Poel (2006) measure number and type of products that each customer owns in each period of time, and total savings and total liabilities in each period of time, respectively.

> *Condition 2*: All customers owning transaction accounts, custody accounts and savings deposits are defined as active customers when these accounts showed a positive balance of at least 50 euros. Intensity of product ownership (measured by average monthly assets and average monthly liabilities) helps us to check this second condition to define a customer as active or inactive.

Finally, the adoption and usage of online banking clearly influences the product choice. This is because the opportunities to use online capabilities to increase sales through add-on sales are enormous (Sarel and Marmorstein, 2003). Banks have a relatively long history of introducing technologies aimed at lowering the costs of customer interaction (e.g., Automated Teller Machines – ATM, centralised telephone call centres, touch-tone banking) and online banking is one of the latest technologies that they have integrated as a new distribution channel (Campbell and Frei, 2010). The Internet has emerged as a key competitive arena for their future because online banking offers customers more features with lower cost than traditional banking activities (Han and Baek, 2004). Internet banking is easier, more convenient and offers more features with lower cost than banking in the eighties or nineties. Additionally, there has been limited attention to how much such technologies alter actual customer demand for services and/or the financial performance of individual relationships (Campbell and Frei, 2010). All these reasons justify the inclusion of the adoption of online banking as a predictor of product choice (called here as ***(adoption of online banking)***$_{i,t}$).

4.2. The second component of CLV: *PRODUCT USAGE*$_{ij,t}$

As the second component of CLV we consider ***product usage***, i.e., the quantity or number of banking products of each type that each customer purchases and owns (depth dimension). In a contractual setting, a customer's usage behaviour is observed every period. This behaviour reflects the underlying commitment of customers, because it would be expected that an individual with higher commitment levels also has higher product usage levels (Ascarza and Hardie, 2012).

The dynamic nature of the customer relationship is especially important in service firms, such as financial services retailers, because customers' service usage levels have a substantial impact on the long-term profitability of the organization (Bolton and Lemon, 1999) and moreover on CLV (Verhoef, 2004). While a number of researchers have explored the problem of modelling churn in a non-contractual setting, little attention has been paid to modelling usage in contractual settings (Ascarza and Hardie, 2012). One example is Bolton and Lemon's (1999) research. They use a Tobit model to estimate usage of television entertainment and cellular phone services. In a similar way, Bolton *et al*. (2000) measure the effect of a loyalty program on future usage of a credit card also using a Tobit model.

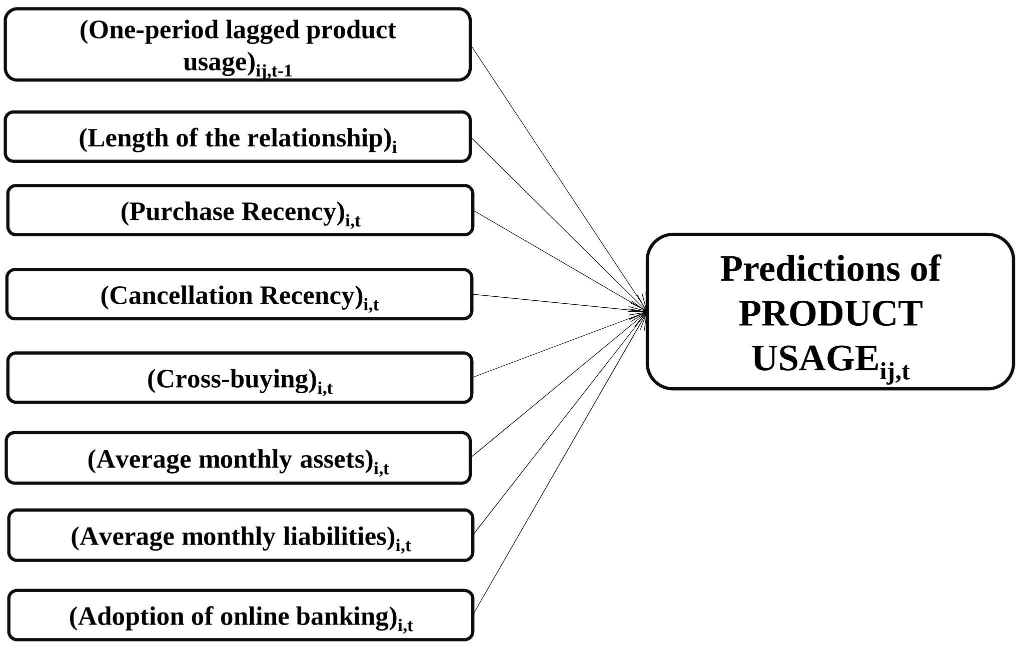
In multi-service industries, customer behaviour is multi-dimensional, that is why the value of a customer of a multiservice provider depends on: (1) not only customer retention or the duration of the provider-customer relationship (length of relationship), but also on (2) the usage level of the consumed services (depth of the relationship), and (3) the number of different services bought from the same provider (breadth of the relationship) (Bolton *et al*. 2004; Verhoef *et al*., 2001). For these companies, customer retention by itself is not fully responsive to the goal of value creation.

Accounting for these behaviours at the individual level (i.e., customer retention, cross-buying and usage) results in more complex, but also more realistic models. Therefore, ***product usage*** is the second dimension of CUSAMS framework (i.e., depth of the relationship) and it is also the second component of CLV.

4.2.1. Drivers of PRODUCT USAGE$_{ij,t}$

Therefore, we model the product choice *j* from a customer *i* in period *t* (in Figure 5 is called ***predictions of PRODUCT USAGE$_{ij,t}$***) using a Poisson distribution (Ascarza and Hardie, 2012). This is because it is a suitable distribution to model count data (for more details see Chapter 5). For this task we have selected several drivers or antecedents of this product usage (and by extension they are also drivers of CLV), namely: previous number of product owned (one-period lagged variable), length of the relationship, recency variables (measured by purchase recency and cancellation recency), cross-buying variable, intensity of product ownership variables (measured by average monthly assets and average monthly liabilities) and adoption of online banking.

Figure 5. Drivers of product usage (*)



(*) where *i* is the customer index, *t* is the time period index, *(t-1)* refers to the previous period of data, and *j* is the banking product index.

The antecedents that we adopt to predict product usage are based on findings from previous research about this same topic (that is, product usage as a component of CLV). The first driver is a one-period lagged variable that measures the number of products owned by each customer (Bolton and Lemon, 1999; Venkatesan *et al*., 2007 p. 585), called here ***(one-period lagged product usage)***$_{ij,t-1}$. The findings of previous authors suggest that customers decide how much to use the service in the future by considering how resources currently are exchanged within the provider-customer relationship. Additionally, and as we have noted previously, the inclusion of lagged values facilitates prediction and interpretation of causality. Therefore, we use the one-period lagged value for one of the antecedents in our analysis (Villas-Boas and Winer, 1999). Length of the relationship and recency variables are two more frequent predictors of product usage (Bolton *et al*., 2000; Venkatesan *et al*., 2007 p. 585, respectively), they are called ***(length of the relationship)***$_i$, ***(purchase recency)***$_i$ and ***(cancellation recency)***$_i$. Product usage has been identified as a consequence of cross-buying (Kumar *et al*., 2008a), or in other words, with an higher number of product categories to choose from (or higher cross-buy), the likelihood of placing an order (more usage) increases. Therefore, inverting this argument, cross-buying can be also considered as a predictor of product usage (called ***(cross-buying)***$_{i,t}$). One of the interesting elements of Reinartz *et al*.'s (2008) research was the analysis of the dispersion (spread or concentration) of spending across product categories. To measure the degree of spread (or concentration) of spending across the different categories, they define a variable called 'balance', which coincides with the intensity of product ownership variable in our study. We also use the balance or intensity of product ownership variable as predictor of product usage (measured by ***(average monthly assets)***$_{i,t}$ and ***(average monthly liabilities)***$_{i,t}$), in particular to distinguish between assets and liabilities of each customer and to see if this distinction influences the product usage. Finally, the adoption and usage of online banking (called ***(adoption of online banking)***$_{i,t}$) also influences the product usage because, on average, PC banking customers use more products than the traditional customer population (Hitt and Frei, 2002).

4.3. The third component of CLV: *CONTRIBUTION MARGIN*$_{i,t}$

The third component of CLV is the ***contribution margin***. For some authors that analyse CLV in the context of retail banking, contribution margin is defined as the revenue resulting from interest payments and commissions fees less liquidity cost, equity cost, risk cost and transaction cost covering the cost of the bank of holding cash, maintaining a certain credit risk-dependent equity ratio, accepting the risk of credit loss and carrying out customer-related transactions respectively (Haenlein *et al*., 2007 p.224). Depending on the data available, researchers use different assumptions regarding the margin or contribution margin. For example, some of them use a margin that is not consumer specific and that is calculated taking into account the defection rates of the customers buying service *j* (Benoit and Van den Poel, 2009). Other authors include only information about revenues because cost was not available (Tirenni *et al*., 2007 p. 135), or compute the profit of each transaction, by a business rule, say a margin of 1% of the amount exchanged at the transaction (Glady *et al*., 2009a). In our research, margin is defined as the difference between interest and fees charged to the customer minus the bank cost paid or the bank incomes earned (because of the bank invests the money of customers funds and other products) at the *Interbank Lending Market*[7].

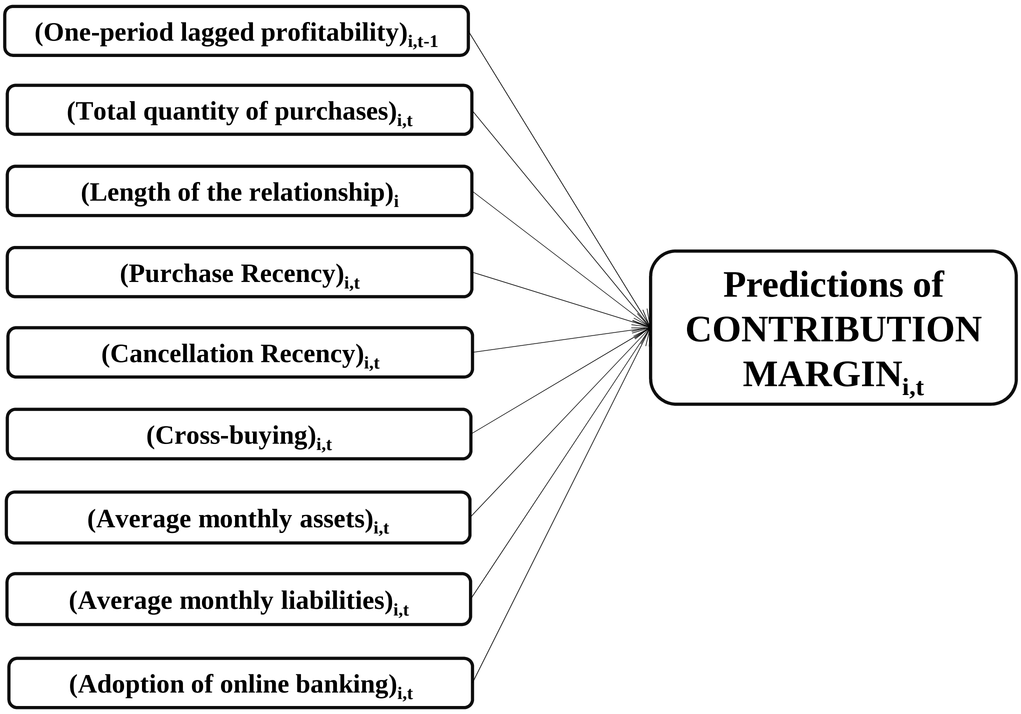
For sakes of simplicity researchers usually assume that the margins of products remain constant over time. Such is the case for Verhoef and Donkers (2001) and Verhoef (2004). However, this assumption is questionable as we show in the following example. If someone that is estimating a CLV model in a particular context, for example a grocery retailer, considers as margin the average margin for grocery products, common sense tells us that probably he or she is not going to get very accurate results, because margin as constant variable is not a good assumption. Therefore, ***contribution margin*** is the third component of CLV.

4.3.1. Drivers of CONTRIBUTION MARGIN$_{i,t}$

Therefore, we model the contribution margin from a customer *i* in period *t* (in Figure 6 it is called ***predictions of CONTRIBUTION MARGIN$_{i,t}$***) using a normal process (Abe, 2009b; Borle *et al*., 2008). This is because contribution margin is a continuous variable (expressed in real numbers, which includes zero, positives and negatives values with decimals) (for more details see Chapter 5). For this task we have selected several drivers or antecedents of this contribution margin (and by extension they are also drivers of CLV). In particular, there are: previous contribution margin or profitability (one-period lagged variable), total quantity of purchases, length of the relationship, recency variables (measured by purchase recency and cancellation recency), cross-buying variable, intensity of product ownership variables (measured by average monthly assets and average monthly liabilities) and adoption of online banking.

[7] The *Interbank Lending Market* is a market in which banks extend loans to one another for a specified term. Most interbank loans are for maturities of one week or less, the majority being overnight. Such loans are made at the interbank rate (also called the overnight rate if the term of the loan is overnight). Low transaction volume in this market was a major contributing factor to the financial crisis of 2007.

Figure 6. Drivers of contribution margin (*)



(*) where *i* is the customer index, *t* is the time period index, (*t-1)* refers to the previous period of data, and *j* is the banking product index.

The antecedents that we adopt to predict contribution margin are based on findings from previous research about this same topic (that is, margin as a component of CLV). The first two drivers that we include as predictors are a one-period lagged variable of profitability (called ***(one-period lagged profitability)****$_{ij,t-1}$*) and the total quantity of purchases (called ***(total quantity of purchases)****$_{i,t}$*). The inclusion of lagged values, as we have noted before, facilitates prediction and interpretation of causality, but also addresses the issue of endogeneity (in a statistical model an endogeneity problem exists when there is a correlation between an independent variable and the error term). Therefore, we use the one-period lagged value for one of the antecedents in our analysis (Villas-Boas and Winer, 1999). For the second driver, total quantity of purchases, it has been verified that both the previous profitability and the total quantity of purchases affect the future contribution margin (Venkatesan and Kumar, 2004). Other authors include length of the relationship (Kumar and Shah, 2009) and also recency variables (Kumar *et al*., 2006a) as predictors of margin (called ***(length of the relationship)****$_i$*, ***(purchase recency)****$_i$* and ***(cancellation recency)****$_i$*, respectively). Cross-buying (called ***(cross-buying)****$_{i,t}$*) is included as predictor of contribution margin as well (Kumar and Shah, 2009). In particular, Kamakura *et al*. (2003 p. 47) and later Prinzie and Van den Poel (2008 p. 714) note that "*cross-selling is effective for customer retention by increasing switching costs and enhancing customer loyalty, thus directly contributing to customer profitability and lifetime value (CLV)*", because cross-buying increases the total value of the customer along his or her relationship with the firm (Kamakura *et al*., 2003). Prinzie and Van den Poel (2006) explain that past and current purchase behaviours are reflected by (current) intensity of product ownership. Therefore, intensity of product ownership or balance (measured by ***(average monthly assets)$_{i,t}$*** and ***(average monthly liabilities)$_{i,t}$***) is also a good predictor of contribution margin and also of CLV (Haenlein *et al*., 2007). Finally, we have found some controversy as to the effect of the usage of online banking on contribution margin. In particular, some authors point out that PC banking customers, on average, offer a higher contribution margin than the traditional customer population (Hitt and Frei, 2002). However, other authors find that the usage of online banking results in an increase in cost-to-serve customers (by the combination of the use of more costly self-service delivery channels, such as Automatic Teller Machines – ATM– or online channels, and a substantial increase in total transaction volume) (Campbell, 2006; Campbell and Frei, 2010). Therefore, the adoption and usage of online banking is configured as another driver of contribution margin (called ***(adoption of online banking)$_{i,t}$***).

4.4. The fourth component of CLV: *DISCOUNT RATE$_t$*

Finally, the last component of CLV is the ***discount rate*** (*d*). To develop a completed CLV model, researchers use past data to predict future estimations, such is the case of contribution margin (measured in euros). Using *d* implies taking into account the time value of money to adjust back the predictions about the future to the present. The value of money is not constant across time and since the money received today is more valuable than the money received in future time periods, these future predictions are discounted to the present value (Kumar, 2008b). Chang (2011) used 'market interest rates' as discount rate. In a financial service setting, the discount rate *d* depends on the general rate of interest and is normally proportional to the treasury bill or the interest that banks pay on savings accounts (Kumar, 2006). It can also vary across firms depending upon the cost of capital to the firm.

4.5. Value based segmentation variables

Prior research incorporates ***socio-demographic variables*** in an attempt to explain CLV (e.g., Abe, 2009b; Kumar *et al*., 2006a; Reinartz and Kumar, 2000; Verhoef and Donkers, 2001). Whereas some studies support this inclusion, others find demographics to be poor predictors of different customer behaviours due to their weak explanatory power, their indirect effect and their association with psychographics (Ailawadi *et al*., 2001a, 2001b). In our model, socio-demographic variables are not used as predictors of customer behaviour, particularly for the critics that they have received, but following the suggestions of Bruhn *et al*. (2006) and other authors, such as Keiningham *et al*., (2006) or Kumar *et al*. (2009), we propose to use them as input of an *ex poste* segmentation. That is, once we have calculated each CLV_i, where i represents an individual customer, we are going to get a customer classification (using data mining techniques, in particular regression trees; for more details see Chapter 5) based on that value and certain socio-demographic variables. Below, we have noted several ideas related to these socio-demographic variables that help us to justify their inclusion as segmentation variables.

Regarding age and income, we refer to the research of Garland (2002, 2004), who indicated that customer contribution (defined as relationship revenue minus relationship cost) is significantly influenced by the customer's age. Garland analysed 1.100 personal retail customers of a bank. Using a stepwise regression analysis, he shows that out of 26 non-financial profitability drivers (related to perceived service quality, customer satisfaction, customer loyalty and customer demographics), only four had significant explanatory power for customer contribution. They are: age, share-of-wallet, household income and joint accounts, with age being the most important one. Additionally, Campbell and Frei (2004) underline that age can be assumed to influence profitability by its impact on consumption patterns in a banking context (e.g., middle-aged customers tend to be more profitable than younger ones because they tend to maintain higher balances and are more likely to have mortgages). They highlight the importance that this type of socio-demographic data has in the day-to-day reality of many U.S. retail banks: *"... A typical retail financial services company spends between $1 million and $2 million annually to procure demographic data from outside vendors"* (Campbell and Frei, 2004 p. 110).

Buckinx and Van den Poel (2005) point out the extensive use of customer demographics in other studies related to customer defection, one of the drivers of CLV. For example, Mittal and Kamakura (2001) show that gender, the number of children in a household, as well as the area of residence are moderating customer characteristics. Vakratsas (1998) confirms the moderating role of household size (i.e., small households are more likely to defect than larger size households). Also, Mozer *et al*. (2000) include an indication of the subscriber's location when they explore techniques from statistical machine learning (i.e., Logit regression, decision trees, neural networks and boosting) to predict churn. Based on these predictions, they determine what incentives should be offered to subscribers to improve retention and maximise profitability to the firm.

Therefore, although demographic variables often tend to be only weak predictors of future behaviour, we decide to include them in our research due to their high relevance in business life. In particular, we include **age, gender** and **income** for the purpose of *ex poste* segmentation.

The main idea of organizing customer characteristics is to target very specific customer groups with controllable variables (Woo *et al*., 2005) and for this task, the customer characteristic should be actionable and differential. In many research studies it is not easy to extract discriminant variables from the target group using only descriptive information such as demographics, geographic bases, and socioeconomics (Drozdenko and Drake, 2002). For this reason we combine socio-demographic variables and the CLV model output as inputs for an *ex poste* segmentation analysis.

Chapter 5. METHODOLOGY

In this chapter we present and describe in detail how we have designed a new approach to calculate CLV and segment customers of a Spanish financial multi-services retailer. In short, we have implemented a two-stage model, in which the first stage implies the development of a stochastic behavioural model to estimate and predict individual CLV's (CLV_i, where i is the customer index), taking into account the behavioural measures that we have explained in the previous Chapter. The second stage implies an *ex poste* segmentation of customers, taking into account the CLV model output and several socio-demographic variables.

5.1. First empirical stage: *COMPUTING CLV FOR EACH CUSTOMER*

We define CLV as: ***CLV as the net present value of the sum of the current and future contribution margins from the customers of the company*** (predictions are estimated over a future time horizon of one year or 12 months), ***which depends on length, depth and breadth of the relationship, over their lifetimes of operation with the company, taking into account the time value of money using a discount rate to adjust back the predictions about the future to the present***. The prediction horizon is held at one year and not the natural lifespan of the customer as the term 'lifetime value' may imply. Some previous studies have used a prediction window of three years (e.g., Kumar *et al*., 2008b; Venkatesan and Kumar, 2004; Venkatesan *et al*., 2007), because given the dynamic environment in which firms typically operate, three years offer a good trade-off between prediction accuracy and prediction horizon when computing the CLV at an individual customer level (Kumar and Shah, 2009 p.123). Furthermore, in general, the concept of discounting future cash flows results in a majority of the customers' lifetime value being captured within the first three years (Gupta and Lehmann 2005). However, using hierarchical Bayesian models, we need the values of the independent variables in order to predict the values of the dependent ones for the next periods. Therefore, as we have data of 24 months of banking operations, we have split the sample into two parts (only in case of the dependent variables), the first 12 months for the dependent variables and the second 12 months also for the dependent variables. We have deleted the last 12 months of the observed dependent variables (i.e., from month number 13 to month number 24) in order to predict them and test the predictive accuracy of the selected technique. To test the previously mentioned predictive accuracy, we have compared the predicted values with the observed ones analytically and graphically (for more details see section 6.4.4.). Other previous authors also used a prediction horizon of only one year. Such is the case of Hwang *et al*. (2004).

Thus, following the suggestions of Donkers *et al*. (2003, 2007) and Verhoef (2004), we can specify the CLV for customer *i* as follows:

$$CLV_i = \sum_{t=1}^{T} \frac{Profit_{i,t}}{(1+d)^t}$$

Where:

CLV_i = lifetime value for customer *i*,

i = index for customers ($1 \leq i \leq I$, *I* is the total sample size),

t = index for periods of time or months ($1 \leq t \leq T$, T is the end of the calibration or observation time frame; we have used all the 24 months of independent variables and the first 12 months of dependent ones in order to predict the last 12 months of the dependent ones),

$Profit_{i,t}$ = current and future (predicted) contribution margins from the customers of the company, and

d = monthly discount factor, which is the fourth component of CLV.

Additionally, we calculate $Profit_{i,t}$, the main input to get CLV_i. This equation contains three terms that must be predicted for each customer (for more details see below). Thus, we can specify profit for customer *i* and period (month) *t* as follows:

$$Profit_{i,t} = \sum_{j=1}^{J} PRODUCT\ OWNERSHIP_{ij,t} * PRODUCT\ USAGE_{ij,t} * CONTRIBUTION\ MARGIN_{i,t}$$

Where:

$Profit_{i,t}$ = current and future (predicted) contribution margins from customers of the company (i) each time period (t, $1 \leq t \leq T$),

j = index for banking products ($1 \leq j \leq J$, J is the total number of products),

$PRODUCT\ OWNERSHIP_{ij,t}$ = observed and predicted values of the first component of CLV,

$PRODUCT\ USAGE_{ij,t}$ = observed and predicted values of the second component of CLV, and

$CONTRIBUTION\ MARGIN_{i,t}$ = observed and predicted values of the third component of CLV.

Finally, we can also calculate CE. Thus, following the suggestions of Rust *et al*. (2000, 2004a), we adapt their formula to our context and data available. CE for the sample of customers is specified as follows:

$$CE = mean(CLV_i) * POP = \sum_{i=1}^{I} CLV_i\,, if\ POP = sample$$

Where:

$mean\ (CLV_i)$ = average lifetime value for firm customers (i) across the sample and,

POP = total number of customers in the sample.

Some authors have given detailed overviews and comparisons of the wide range of different approaches that have been used for CLV modelling (e.g., Donkers *et al*., 2007; Gupta *et al*., 2006; Kumar and George, 2007; Ngai *et al*., 2009). In particular, Donkers *et al*. (2007) explained that ***regression type models*** are often used in this context, e.g., linear regression model (Malthouse and Blattberg, 2005; Malthouse and Mulhern, 2008); the Probit model (Verhoef and Donkers, 2001); the multivariate Probit model (Donkers *et al*., 2007); the multivariate Logit model (Prinzie and Van den Poel, 2007). These types of models have the disadvantage that they are smoothing techniques that attempt to describe well the relationship between the predictors and the response but tend to treat heterogeneity as noise (Colombo and Jiang, 1999). Moreover, CLV has been analysed in a substantial number of different research domains, ***varying from econometric models to computer science techniques*** (Gupta *et al*., 2006).

It is noteworthy that research on CLV measurement has so far focused on specific contexts (ad hoc), because the data available to a researcher or firm in different contexts might be different. The two types of context generally considered are: non-contractual and contractual (e.g., Reinartz and Kumar 2000, 2003; for more details see section 3.1.1). Different models for measuring CLV arrive differently at estimates of the expectations of future customer purchase behaviour. A popular method that follows such an approach in a non-contractual context is the ***negative binomial distribution (NBD)-Pareto model*** by Schmittlein *et al*. (1987). In this model, past customer purchase behaviour (measure of purchase frequency and amount spent during a purchase) is used to predict the future probability of a customer remaining in business with the firm (the probability of each customer being 'alive'). This probability can be used to estimate CLV (Reinartz and Kumar 2000, 2003; Schmittlein and Peterson, 1994). The NBD-Pareto model is applied in instances where customer lifetimes are not known with certainty (i.e., it is not known when a customer stops doing business with a firm). The model assumes that individual customer lifetimes with the firm are exponentially distributed. But, as was discussed by Schmittlein and Peterson (1994), in contexts (such as ours), where customer lifetimes are observed (i.e., the analyst has data on customer lifetime or duration of the relationship), the NBD-Pareto model has limitations and is not suitable (for more details see Customer Lifetime Value (CLV) section in Chapter 2). As we have noted in Chapter 3, section 3.3.1, a contractual setting is the best way to tackle our problem.

The strategic importance of customer assets to create a competitive advantage demands high quality models that predict the CLV with as less error as possible. Until now, most models only look at the absolute value of the future cash flows without considering uncertainty of these values (Haentjens, 2011). To solve this task, some authors have proposed to use a ***Bayesian approach*** (Rossi and Allenby, 2003) because it can naturally incorporate past behavioural outcomes into future expectations. The Bayesian approach is defined as "*the explicit use of external evidence in the design, monitoring, analysis, interpretation and reporting of a scientific investigation*" (Spiegelhalter, 2004). Therefore, Bayesian decision theory postulates that there are uncertain states (e.g., quantity and timing of purchases by customers) as a result of actions that a firm can take (e.g., setting marketing decision variables, selecting customers). Typical sources of this uncertainty are related to the poor or non-existent information in most CRM databases on customer transactions with the competition, competitor marketing actions targeted to each customer and customer attitudes. The combination of states and actions results in consequences for the firm (e.g., profits). Bayes' theorem combines data with prior distributions for these states to obtain posterior distributions to reduce the uncertainty about the states and at the same time to reduce the possibility of errors when making long-term predictions (Venkatesan *et al*., 2007).

Jain and Singh (2002) call for more research in accurately predicting CLV based on the history of usage and prior estimates of CLV, for instance using the previously mentioned Bayesian approach, and providing more accurate estimates of CLV than the traditional regression analysis of historical data. Bayesian methods can incorporate such prior information in the structure of the model easily through the priors of the distributions of the drivers of CLV (Abe, 2009b; Borle *et al*, 2008). Furthermore, this approach can be used in any context (that is, contractual or non-contractual). In particular, Borle *et al*. (2008) and Abe (2009b) call for the inclusion of a rich set of covariates in their Hierarchical Bayesian framework to estimate CLV. Our drivers and components, mentioned previously, can enrich the estimation of CLV through the Hierarchical Bayes approach and at the same time we can prove if they have a real effect on the value of each customer (for more details about HB to develop a CLV model, see some of the most important empirical illustrations such as Venkatesan *et al*., 2007; Borle *et al*., 2008; Abe, 2009b; Kumar and Shah, 2009).

Therefore, for the accurate measurement of CLV_i, we have used Bayesian analysis to estimate its three core components (i.e., $PRODUCT\ OWNERSHIP_{ij,t}$, $PRODUCT\ USAGE_{ij,t}$, and $CONTRIBUTION\ MARGIN_{i,t}$), using suitable distributions and then combine the predictions from these three models to arrive at a single value representing CLV_i in terms of euros. In particular, a hierarchical Bayesian model estimation procedure is adopted to estimate the parameter of interest, which are the regression coefficients associated with the three parts of the models and the predictions. In other words, we use a hierarchical Bayesian model to test if the selected covariates have the potential to predict the three core components of CLV and finally, predict these three core components of CLV. The hierarchical nature of the model is reflected by the fact that the parameters (the priors) of the three distributions (product ownership~Bernoulli model: p_{it}; product usage~Poisson model: λ_{it}; and contribution margin~normal model: μ_{it}) are expressed as a function of the available covariates of the customer (length of the relationship, purchase recency, cancellation recency, cross-buying, average monthly assets, average monthly liabilities, adoption of online banking, total quantity of purchases and several one-period lagged variables).

5.1.1. Bayesian modelling

A ***probabilistic model*** represents, or sufficiently approximates, the true generating mechanism of a phenomenon under study. Probabilistic and logical arguments guide the construction of any probabilistic model and its *likelihood* (or joint distribution) contains the available information provided by the observed sample, as is specified below:

$$f(y|\theta) = \prod_{i=1}^{n} f(y_i|\theta)$$

Where Y is a random variable called response, which follows a probabilistic rule with density or probability function previously shown $f(y|\theta)$, where θ is the parameter vector and $y = [y_1, \ldots, y_n]^T$ is an independent, identically distributed (i.i.d.) sample of size *n* of this variable.

Usually, models are constructed to assess or interpret causal relationships between the response variable Y and several covariates or explanatory variables called X_j. In such cases, X_j are linked with the response variables via a deterministic function and part of the original parameter vector is substituted by an alternative set of parameters (called β) that usually encapsulate the effect of each X_j on Y. However, ***Bayesian analysis*** differs from the classical statistical theory since all unknown parameters are considered as random variables. For this reason, the ***prior distribution*** must be defined initially. This prior distribution expresses the information available to the researcher before any data are involved in the statistical analysis. Interest lies in the calculation of the ***posterior distribution*** $f(\theta|y)$ of the parameter θ given the observed data y. According to the ***Bayes theorem***[8], the posterior distribution can be written as:

$$f(\theta|y) = \frac{f(y|\theta)f(\theta)}{f(y)} \propto f(y|\theta)f(\theta)$$

[8] Let us consider two possible outcomes A and B. Moreover, assume that $A = A_1 \cup \ldots \cup A_n$ for which $A_i \cap A_j = \emptyset$ for every $i \neq j$. ***Bayes' theorem*** provides an expression for the conditional probability of A_i given B, which is equal to:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

For any outcome A and B, ***Bayes' rule*** provides a simpler and more general form (Hoffmann-Jørgensen, 1994 p.102):

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \propto P(B|A)P(A)$$

The posterior distribution, which is the key element in Bayesian inference, embodies both prior and observed data information, which is expressed by the prior distribution $f(\theta)$ and the same likelihood that we have specified at the beginning of this section (Ntzoufras, 2009 p. 3).

In order to complete the definition of a Bayesian model, both the prior distribution and the likelihood must be fully specified. Specification of the prior distribution is important in Bayesian inference since it influences the posterior inference. Usually, specification of the prior mean and variance is emphasised. The prior mean provides a prior point estimate for the parameter of interest, while the variance expresses our uncertainty concerning this estimate. When we a priori strongly believe that this estimate is accurate, then the variance must be set low, while ignorance or great uncertainty concerning the prior mean can be expressed by large variance. If prior information is available, it should be appropriately summarised by the prior distribution. This procedure is called ***elicitation*** of prior knowledge. Very often, no prior information is available. In this case we need to specify a prior that will not influence the posterior distribution and "*let the data speak for themselves*". Such distributions are frequently called ***non-informative*** or ***vague prior distributions*** (Ntzoufras, 2009 p. 5).

5.1.1.1. Definition of a Bayesian model

In particular, and following the suggestions of Ntzoufras (2009), in order to complete the definition of a Bayesian model, we may divide the whole procedure into five stages: (a) model building; (b) calculation of the posterior distribution; (c) analysis of the posterior distribution; (d) diagnostic tests concerning the appropriateness of the adopted model and the robustness of the posterior distribution; and (e) inference and predictions.

a) Stage 1: Model building

This first stage implies a number of different tasks, such as identify the response variable/s (*Y*) and the corresponding data y (in our case, product ownership, product usage and contribution margin), find a distribution that adequately describes *Y*, identify explanatory variables, build a structure for the parameters of the distribution (using deterministic functions) and write down the likelihood of the model.

To specify the likelihood using WinBUGS, we have to assume a response variable Y with n observed values stored in a vector y with elements y_i. The stochastic part of the model can be written as:

$$Y \sim Distribution\ (\vartheta)$$

Where ϑ is the parameter vector of assumed distribution. The parameter vector is 'linked' with some explanatory variables X_1, X_2, …, X_p using a link function h, as follows:

$$\vartheta = h(\theta, X_1, X_2, \dots, X_p),$$

Where θ is a constrained set of parameters used to specify the link function and the final structure of the model. The vector ϑ is the actual set of parameters to be estimated. Moreover, each subject's specific observations of covariates x_{1i}, x_{2i}, …, x_{pi} will define a different set of parameters θ for each subject *i* given by:

$$\vartheta_{(i)} = h(\theta, x_{1i}, x_{2i}, \dots, x_{pi}),$$

In Generalized Linear Models (GLM)[9], this distribution associates (or links) the parameters of the assumed distribution with a linear combination of the explanatory variables. The likelihood of the model is given by:

$$f(y|\theta) = \prod_{i=1}^{nN} f\left(y_i \middle| \vartheta_{(i)} = h(\theta, x_{1i}, x_{2i}, \dots, x_{pi})\right)$$

The corresponding WinBUGS sintax is given by:

```
for (i in 1:n) {

    y ~ distribution.name (parameter1[i], parameter2[i], …)

    parameter1[i]<-[function of theta and X's]

    parameter2[i] <-[function of theta and X's]

    … } #this is a comment in WinBUGS
```

[9] GLM are a wide class of statistical models encompassing stochastic representations used for the analysis of both quantitative and qualitative response variables. They can be regarded as the natural extension of normal linear regression models and are based on the exponential family of distributions, which includes the most common distributions such as the normal, binomial and Poisson.Three are three components of a GLM: (i) the random/stochastic component (which contains the response variable *Y*), (ii) the systematic component or linear predictor (it is a function of the explanatory variables or covariates *x*) and (iii) the link function (it is the mathematical expression which connects the parameter of the response *Y* with the linear predictor and the covariates *x*) (Ntzoufras, 2009 p.229).

To complete the Bayesian model specification, we further need to specify the prior distribution of the model parameters θ. Thus, we complete the specification by writing:

```
theta1 ~ distribution.name(…)

theta2 ~ distribution.name(…)
```

Some of the most common prior distributions that the researchers usually assume are called *non-informative priors*. This kind of prior distributions are assumed when nothing is known about the value of a parameters. This is a rectangular distribution over the feasible set of values of the parameter. From the normal distribution, an important special case to represent ignorance is `dnorm(0,`ε`)`, where ε is a small number such as 0,001 (in WinBUGS for the special case of a normal distribution: $\varepsilon = \frac{1}{\sigma^2} = precision$). From the gamma distribution to represent ignorance we can use `dgamma(`ε, ε`)`, where ε is also a small number such as 0,001.

b) Stage 2: Calculation of the posterior distribution

To calculate the posterior distribution, we have to identify the method of calculation first (analytically, asymptotically or using simulation techniques) and then implement the selected method to estimate it. After we have identified the simulation method, we need to specify some initial values in the WinBUGS code in order to initiate the MCMC sampler. These initial values must be provided for all stochastic nodes (random variables of the model that are characterised by a distribution, that is, the model parameters (θ)) except for response data/variables. For this research, we use a simulation technique to get samples from the posterior distribution, namely the Markov chain Monte Carlo method (MCMC). In general, simulation methods have solved computational problems and specifically MCMC are very well suited to estimate models that are built from a sequence of conditional distributions, called hierarchical models.

More specifically, WinBUGS (BUGS = Bayesian inference Using Gibbs Sampling) uses the Gibbs sampling algorithm (Geman and Geman, 1984) as a MCMC method[10]. This simulation technique used by the software WinBUGS enables quantitative researchers to use highly complicated models and estimate the corresponding posterior distribution with accuracy.

c) *Stage 3: Analysis of the posterior distribution*

We can analyse the posterior distribution through the visual inspection of the marginal posterior distributions of interest (e.g., using different types of plots, such as marginal posterior density or probability plots, marginal posterior histograms for continuous variables and bar charts for discrete or categorical variables, boxplots, etc.). We can also calculate posterior summaries (such as means, medians, standard deviations, correlations, quantiles), 95% or 99% posterior credible intervals, mode and area of highest posterior density (where possible).

d) *Stage 4: Diagnostic tests concerning the appropriateness of the adopted model and the robustness of the posterior distribution (via sensitivity analysis)*

As we generate posterior distribution using MCMC, we have to apply diagnostic tests to monitor the convergence of the MCMC algorithm (with the term convergence of the algorithm, we refer to situations where the algorithm has reached its equilibrium and generates values from the desired target distribution). Additionally, the Monte Carlo error is an important measure that must be reported and monitored. It measures the variability of each estimate due to the simulation. MC error must be low in order to calculate the parameter of interest with increased precision.

e) *Stage 5: Inference and predictions*

Bayesian theory provides a realistic and straightforward theoretical frame for the prediction of future observations through the *predictive distribution* (Ntzoufras, 2009 p. 341). It is equivalent to the fitted (or expected or predicted) values in classical theory with the difference that now we

[10] Cowles (2004) indicates that WinBUGS uses the Gibbs sampling algorithm to construct the transition kernels for its Markov chain samplers. Each iteration of a Gibbs sampler involves drawing a new value for each parameter from its 'full conditional distribution', i.e., the conditional probability distribution of that parameter given the current values of all other quantities in the model. During compilation, WinBUGS chooses a method to draw samples from each the full conditional distribution of each model parameter. These methods are: the slice-sampling algorithm (Neal, 1997) and the random walk Metropolis algorithm (Metropolis et al., 1953). Samples from the tuning or burn-in phases of both the slice-sampling and Metropolis algorithms are ignored in the calculation of all summary statistics, although they will appear in trace plots (Cowles, 2004 p.335).

directly deal with a distribution. This predictive distribution is the distribution of the data averaged over all possible parameter values. For this reason, when data *y* have not been observed yet, predictions are based on the marginal likelihood, which is the likelihood averaged over all parameter values supported by our prior beliefs:

$$f(y) = \int f(y|\theta)f(\theta)d\theta$$

Hence, $f(y)$ is also called *prior predictive distribution*.

Usually, after having observed data *y*, one finds the prediction of future data *y'* more interesting. Following this logic, we calculate the posterior predictive distribution. This distribution is termed as a predictive distribution since prediction is usually attempted only after observation of a set of data *y*. Future observations *y'* can be alternatively viewed as additional parameters under estimation. From this perspective, the joint posterior distribution is now given by: $f(y', \theta|y)$. Inference on future observations *y'* can be based on the marginal posterior distribution $f(y'|y)$ by integrating out all nuisance parameters, one of which in this case, is the parameter vector θ. Hence, the predictive distribution is given by the likelihood of future data averaged over the posterior distribution $f(\theta|y)$:

$$f(y'|y) = \int f(y', \theta|y)d\theta = \int f(y'|\theta, y)f(\theta|y)d\theta = \int f(y'|\theta)f(\theta|y)d\theta$$

This distribution is also used for checking the assumptions of the model proposed and also for fitting the model.

To estimate the predictive distribution for future observations using MCMC, let us to consider a usual normal regression model and an unknown future observation Y_{n+1} with known covariate values $x_{(n+1)} = \left(x_{n+1,1}, x_{n+1,2}, \dots, x_{n+1,p}\right)$ (Ntzoufras, 2009 p. 344). Then we can estimate its expected value $E\left(Y_{n+1}\middle|y, x_{(n+1)}\right)$ using the predictive distribution:

$$f(y_{n+1}|y, x_{(n+1)}) = \int f(y_{n+1}|\beta, \sigma^2, x_{(n+1)})f(\beta, \sigma^2|y)d\beta d\sigma^2$$

Quantity y_{n+1} can be considered as an additional parameter under estimation. Thus, it can be generated within an MCMC scheme from the conditional posterior distribution, since Y_i are independent, identically distributed (i.i.d.) random variables:

$$f\left(y_{n+1}\middle|\beta,\sigma^2,y,x_{(n+1)}\right)=f(y_{n+1}|\beta,\sigma^2,x_{(n+1)})$$

Hence, we only need to generate y_{n+1} from the distribution assumed by its model structure with the appropriate parameter values. For the usual normal regression model, we can generate y_{n+1} in the *t*th iteration of the algorithm by:

$$y_{n+1}^{(t)}\sim N\left(\mu_{n+1}^{(t)},\sigma^{2,(t)}\right) with\ \mu_{n+1}^{(t)}=E\left(Y_{n+1}\middle|\mu_{n+1}^{(t)},x_{(n+1)}\right)=\beta_0^{(t)}+\sum_{j=1}^{p}\beta_j x_{n+1,j}^{(t)}$$

In WinBUGS, we only need to define an additional stochastic node called `ynew` (where `tau` is equal to precision or $\frac{1}{\sigma^2}$):

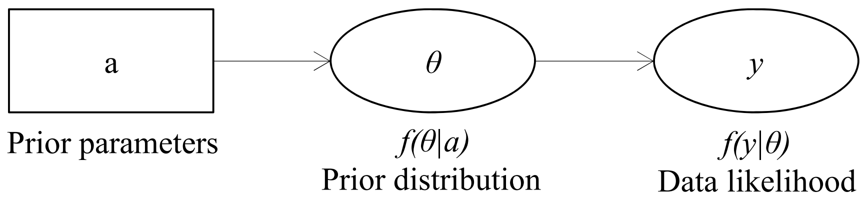
```
ynew ~ dnorm (munew,tau)

munew <- beta0 + inprod(beta[], xnew[])
```

Where `xnew[]` is the vector with elements of the explanatory values for the future (to-be-estimated) response. To complete the specification of the additional nodes, we need to specify `xnew` in the data of the WinBUGS model code. Moreover, in the data section, we must specify that the value of `ynew` is not available by setting `ynew=NA` in the list data format. As we have already mentioned, `ynew` is treated in a manner that is similar to that used for parameters. A simple monitoring of the posterior distribution of `ynew` produces a sample from the posterior predictive distribution, enabling us to calculate posterior summaries, density plots and other properties.

5.1.1.2. Hierarchical Bayesian models

Bayesian models have an inherently hierarchical structure. The prior distribution $f(\theta|a)$ of the model parameters θ with prior parameters a can be considered as one level of hierarchy, with the likelihood as the final stage of a Bayesian model resulting in the posterior distribution $f(\theta|y)\propto f(y|\theta)f(\theta;a)$ via the Bayes' theorem. See Figure 7 for a graphical representation of the hierarchical structure of a typical Bayesian model.

Figure 7. Graphical representation of the hierarchical structure of a standard Bayesian model (*)



(*) Squared nodes refer to constant parameters; oval nodes refer to stochastic components of the model.

Robert (2007) provides a series of justifications and advantages for using hierarchical models, including the fact that the prior is decomposed into two main parts: one referring to structural information or assumptions concerning the model and one referring to the actual subjective information of the model parameters. Another advantage is that hierarchical structure leads to a more robust analysis, reducing subjectivism since posterior results are averaged across different prior choices of parameters of interest. Finally, the hierarchical structure simplifies both the interpretation and the computation of the model since the corresponding posterior distribution is simplified, resulting in conditional distributions of simpler form.

As we have noted at the beginning of this section, the proposed model is a mixture of Bernoulli, Poisson and normal distributions, which are used to jointly estimate the ownership, the usage of banking products and the contribution margin. The hierarchical nature of the model is reflected by the fact that the parameters (the priors) of the three distributions (product ownership~Bernoulli model: p_{it}; product usage~Poisson model: λ_{it}; and contribution margin~normal model: μ_{it}) are expressed as a function of the available covariates of the customer.

5.1.2. Predictions of PRODUCT OWNERSHIP$_{ij,t}$

We consider a set of 18 binary data because the response variable for all 18 products in the database indicates with a 1 if a customer owns a product or 0 otherwise. This response variable is called *product ownership*: *O_01[i,t] ... O_18[i,t]*.

Let p_i be the probability that a customer owns banking products, then to get ***predictions of PRODUCT OWNERSHIP$_{ij,t}$*** (length and breadth dimensions), we use a ***Bernoulli distribution*** (Ntzoufras, 2009 p. 257, 263), thus $O_i \rightarrow Bernoulli\ (p_i)$. The most popular model in this case is the logistic regression model, in which the usual logit link[11] is adopted (Ntzoufras, 2009 p.255). The logit link is not only the obvious choice since it is a canonical link but also has a smooth and nice interpretation based on the odds ratio $p/(1-p)$, which is denoted the odds of $O = 1$ and $O = 0$, where p is the probability of success for O. The idea behind the logit link function is to map an unrestricted covariate space into the restricted parameter space (0,1) of probability p_i. The odds prediction equation is $odds = e^{a+bx}$ and it can be converted into probabilities as follows:

$$\hat{O} = \frac{odds}{1 + odds}$$

We use varying p_i's for each customer to emphasize the differences between customers due to different characteristics, i.e., where $i = 1, \ldots, n$, X_k is a vector of k covariates for customer i, and $\theta = (\theta_0, \theta_1, \theta_2, \ldots \theta_k)$ are the regression coefficients, with θ_0 being the constant term.

The logistic regression model can be summarised by:

$$O_i \sim Bernoulli(p_i),$$

$$log\frac{p_i}{1-p_i} = \theta_0 + \sum_{k=1}^{k} \theta_k x_k = X_k\theta$$

The coefficients θ measure the partial impact of each covariate (x_k) on $\log(\frac{p_i}{1-p_i})$, and consequently, e^{θ} measures the impact on the odds ratio.

Thus, we specify *PRODUCT OWNERSHIP*$_{ij,t}$ for each customer *i*, each product *j* and each month *t* as follows:

$$O_{ij,t} = f\left(L_i, PR_{i,t}, CR_{i,t}, CB_{i,t}, AMA_{i,t}, AML_{i,t}, OL_{i,t}\right), if\ t = 1$$

$$O_{ij,t} = f\left(L_i, PR_{i,t}, CR_{i,t}, CB_{i,t}, AMA_{i,t}, AML_{i,t}, OL_{i,t}, O_{ij,t-1}\right), if\ t > 1$$

Where:

$O_{ij,t}$ = product ownership, where $i = 1, \ldots, 1.357$, $j = 1, \ldots, 18$ and $t = 1, \ldots, 24$,

[11] The link function is a monotonic and differentiable function used to match the parameters of the response variable with the systematic component, namely, the linear predictor and the associated covariates.

L_i = length of the relationship,

$PR_{i,t}$ = purchase recency,

$CR_{i,t}$ = cancellation recency,

$CB_{i,t}$ = cross-buying,

$AMA_{i,t}$ = average monthly assets (intensity of product ownership related to assets),

$AML_{i,t}$ = average monthly liabilities (intensity of product ownership related to liabilities),

$OL_{i,t}$ = adoption of online banking, and

$O_{ij,t-1}$ = one-period lagged variable of product ownership.

5.1.3. Predictions of PRODUCT USAGE$_{ij,t}$

Regarding product usage, the response variable for all products in the database is defined in ℕ, because such variables express the number of successes (total number of products that each customer owns) within a fixed time interval. This response variable is called *product usage*: *U_01[i,t] ... U_18[i,t]*.

To get ***predictions of PRODUCT USAGE$_{ij,t}$*** (depth dimension), we focus on ***Poisson regression models*** (they are frequently called *Poisson log-linear models*). The Poisson distribution is a suitable distribution to model count data (Ascarza and Hardie, 2012).

The random variable U_i then follows a Poisson distribution with parameter λ_i, where λ_i is the expected number of products used by customer *i*:

$$P(U_i = u) = \frac{e^{-\lambda}}{x!}\lambda^u, with\ i = 1,2, \dots, n$$

$$E(U_i) = Var(U_i) = \lambda_i$$

Heterogeneity between customers can be accommodated into the model by assuming λ_i to be a random variable that is influenced by covariates or characteristics of a particular customer, i.e.,

$\lambda_i = e^{X_k\beta}$ or $log\lambda_i = X_k\beta$, where X_k is a vector of k specific covariates, and $\beta = (\beta_0, \beta_1, \beta_2, \dots \beta_k)$ are the regression coefficients, with β_0 being the constant term. The exponent, in the previous equation, guarantees that the predicted product usage is positive.

The Poisson log-linear model is summarised by the following expression:

$$U_i \sim Poisson(\lambda_i),$$

$$log\ \lambda_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_k = X_k\beta$$

The coefficients β measure the partial impact of each covariate (x_k) on $\log(\lambda_i)$, and consequently, e^{β} measures the impact on λ_i.

Thus, we specify *PRODUCT USAGE*$_{ij,t}$ for each customer *i*, each product *j* and each month *t* as follows:

$$U_{ij,t} = f\big(L_i, PR_{i,t}, CR_{i,t}, CB_{i,t}, AMA_{i,t}, AML_{i,t}, OL_{i,t}\big), if\ t = 1$$

$$U_{ij,t} = f\big(L_i, PR_{i,t}, CR_{i,t}, CB_{i,t}, AMA_{i,t}, AML_{i,t}, OL_{i,t}, U_{ij,t-1}\big), if\ t > 1$$

Where:

$U_{ij,t}$ = product usage, where *i = 1,…, 1.357*, *j = 1,…, 18* and *t = 1,…, 24*,

L_i = length of the relationship,

$PR_{i,t}$ = purchase recency,

$CR_{i,t}$ = cancellation recency,

$CB_{i,t}$ = cross-buying,

$AMA_{i,t}$ = average monthly assets,

$AML_{i,t}$ = average monthly liabilities,

$OL_{i,t}$ = adoption of online banking, and

$U_{ij,t-1}$ = one-period lagged variable of product usage.

5.1.4. Predictions of CONTRIBUTION MARGIN$_{j,t}$

Finally, to get ***predictions of CONTRIBUTION MARGIN$_{i,t}$***, we use a **normal process** (Abe, 2009b; Borle *et al*., 2008). For sakes of simplicity researchers usually assume that the margins of products remain constant over time. Such is the case with Verhoef and Donkers (2001) and Verhoef (2004). However, this assumption is questionable and for this reason we have also decided to predict this third component of customer value, the contribution margin. The main reason that justifies the choice of a normal process is that our dependent variable is defined in ℝ, that is, it is a continuous variable that includes zero, positives and negatives values with decimals (called *contribution margin: CM[i,t]*).

The normal model (Abe, 2009b; Borle *et al*., 2008) can be summarised by:

$$CM \sim normal(\mu_i, \tau_i) \ ,$$

$$\mu_i = \rho_0 + \sum_{k=1}^{K} \rho_k x_k = X_k \rho$$

We use varying μ_i's for each customer to emphasize the differences between customers due to different characteristics, i.e., where $i = 1, \ldots, n$, X_k is a vector of k covariates for customer i, and $\rho = (\rho_0, \rho_1, \rho_2, \ldots \rho_k)$ are the regression coefficients, with ρ_0 being the constant term.

It is important to understand that WinBUGS specifies the normal distribution in terms of the mean (μ) and precision (τ), that is $N(\mu, \tau)$, rather in terms of mean and standard deviation, that is $N(\mu, \sigma)$. The relationship between standard deviation and precision is $\sigma = {}^{1}\!/_{\sqrt{\tau}}$.

Thus, we can specify *predictions of CONTRIBUTION MARGIN$_{i,t}$* for customer i and period (month) t as follows:

$$CM_{i,t} = f\big(L_i, PR_{i,t}, CR_{i,t}, CB_{i,t}, AMA_{i,t}, AML_{i,t}, OL_{i,t}, Q_{i,t}\big), if\ t = 1$$

$$CM_{i,t} = f\big(L_i, PR_{i,t}, CR_{i,t}, CB_{i,t}, AMA_{i,t}, AML_{i,t}, OL_{i,t}, Q_{i,t}, CM_{i,t-1}\big), if\ t > 1$$

Where:

$CM_{i,t}$ = contribution margin, where *i = 1,…, 1.357* and *t = 1,…, 24*,

L_i = length of the relationship,

$PR_{i,t}$ = purchase recency,

$CR_{i,t}$ = cancellation recency,

$CB_{i,t}$ = cross-buying,

$AMA_{i,t}$ = average monthly assets,

$AML_{i,t}$ = average monthly liabilities,

$OL_{i,t}$ = adoption of online banking,

$Q_{i,t}$ = total quantity of purchases, and

$CM_{i,t-1}$ = one-period lagged variable of contribution margin.

5.1.5. Bayesian estimation approach

A hierarchical Bayes model estimation procedure was adopted to estimate the parameters of interest which are the regression coefficients associated with each model. In particular we refer to:

- Product ownership model: $\theta_0, \dots, \theta_8$.
- Product usage model: $\beta_0, \dots, \beta_8$.
- Contribution margin model: $\rho_0, \dots, \rho_9, \tau, \sigma$.

Since we have cases of simple GLM in the exponential family, we follow the standard approaches based on non-informative prior densities for the previously shown regression coefficients and intercepts. Namely, we use normal distributions with zero means and large variances (for $\theta_0, \dots, \theta_8, \beta_0, \dots, \beta_8, \rho_0, \dots, \rho_9$) and gamma distributions with also large parameters (for τ, σ) in order to represent the ignorance about the parameters: $\mathrm{Normal}(0,0.000001)$ and $\mathrm{Gamma}(0.01,0.01)$. Furthermore, in order to test model convergence, two separate chains were run with different starting values for the priors (for more details see Chapter 6). In particular, the first chain was initialized with the values described below. The second group of initial values for the second chain was randomly drawn from the posterior distribution using the command `gen inits` in WinBUGS for product ownership and product usage models; in case of contribution margin the second chain was initialized again using the values described below.

```
#Initial values for the product ownership models:
INITS #for the first chain
list(theta0=1,theta1=1,theta2=1,theta3=1,theta4=1,theta5=1,th
eta6=1,theta7=1,theta8=1)
#second group of initial values randomly drawn
#from the prior distribution –in WinBUGS: geninits-

#Initial values for the product usage models:
INITS #for the first chain
list(beta0=1,beta1=1,beta2=1,beta3=1,beta4=1,beta5=1,beta6=1,
beta7=1,beta8=1)
#second group of initial values randomly drawn
#from the prior distribution –in WinBUGS: geninits-

#Initial values for the contribution margin model:
INITS
#for the first chain:
list(rho0=1,tau=1,rho=c(1,1,1,1,1,1,1,1,1))
#for the second chain:
list(rho0=1,tau=0.5,rho=c(1,1,1,1,1,1,1,1,1))
#WinBUGS: geninits for sigma
```

5.2. Second empirical stage: *VALUE BASED EX POSTE SEGMENTATION*

The second stage of the analysis implies an *ex poste* segmentation of customers, taking into account CLV model output (CLV_i, where i refers to each customer) and several socio-demographic variables (age_i, $gender_i$ and $income_i$). In next sections we review different approaches to perform a value-based segmentation. These approaches are divided into two options: (1) option 1 implies: firstly, segment customers and secondly, get an aggregated CLV measure, and (2) option 2 implies: firstly, get an individual CLV_i (where *i* refers to each customer) and secondly, segment the customer base taking into account these individual CLV's. Finally, we explain and justify the selected value-based segmentation for this research.

5.2.1. Different approaches to perform a value-based segmentation

Option 1: Several authors propose different approaches firstly to segment customers and secondly to get a CLV measure for each group. They use an aggregate view of CLV (e.g., Haenlein *et al*., 2007; Chan, 2008).

These models only predict the average CLV at an aggregate level for the entire customer base, without taking into account characteristics of the individual customer. This is a serious drawback, since profitability is usually not distributed uniformly among customers and a primary objective of the lifetime value approach is to identify highly profitable customers in order to keep existing ones and attract others, and also to identify non-profitable customers in order to reduce or even avoid investment in them (Tirenni *et al*., 2007). This aggregation procedure is also used by, for example Haenlein *et al*. (2007), who firstly use decision trees to segment the customer base, based on certain drivers of profitability, including age, demographics/lifestyle data, product ownership (measured by two variables: type and intensity of product ownership; the first is measured through 11 dummy variables, each one related with each product, where 0 means no ownership of the product and 1 the opposite, and the second is measured by several variables that represent the balances for each customer for each product) and activity level (measured as a dummy variable, where 0 means customer inactive and 1 customer activity with the company). After they segment the customer base, these authors calculate an average CLV for each of the segments obtained. It is quite possible that they could have obtained more accurate CLV results if they had firstly estimated CLV for individual customers and subsequently had segmented the customer base based on CLV output. More recently, Chan (2008) identifies customer behaviour using RFM variables for a Nissan automobile retailer to segment customer base through genetic algorithm, and then uses the CLV model to assess the proposed segmentation.

Option 2: Other authors point out that individual CLV calculation can be used as an intermediate step for classification purposes (e.g., Bruhn *et al*., 2006; Keiningham *et al*., 2006; Kumar *et al*., 2009).

The key for this second option is that CLV measure is an interesting input to perform customer segmentation. While traditional segmentation is focused on identifying customer groups based on demographics and attributes such as attitude and psychological profiles, CLV undertakes a value-based approach that looks at groups of customers in terms of the revenue they generate and the costs of establishing and maintaining relationships with them (Kumar, 2008b p. 43). Therefore, based on the distribution of CLV's, some authors divide, select and prioritise customers into different profitability segments, for example high CLV, medium CLV, and low CLV. Moreover, we have found in the literature several segmentation strategies based on CLV. They can be classified into three categories (Kim *et al*., 2006):

(i) Segmentation by using only CLV values, e.g., the customer pyramid (Zeithaml *et al*., 2001).

(ii) Segmentation by using only CLV components, e.g., current value, potential value, loyalty, etc. (Hwang *et al*., 2004).

(iii) Segmentation by considering both CLV values and other information, e.g., socio-demographic information, transaction history, etc. (Kim *et al*., 2006).

Some authors propose the idea of segmenting the customer base taking into account customer value for this task (Bruhn *et al*., 2006; Lemon and Mark, 2006). Recently, Kumar *et al*. (2009) propose to reverse the conventional path to profitability, firstly examining profitability of each customer (considering their loyalty and satisfaction), which can be efficiently measured by the Customer Lifetime Value (CLV), and secondly, using these results to proceed as the starting point for all management decisions of the customer relationship. This means that when firms adopt an approach based on CLV they are more able to make consistent decisions on how to acquire and retain customers and to identify customers that are not interesting enough in investing in, that is, it enables the company to decide more effectively on the allocation of resources to each customer segment (Kumar and Rajan, 2009).

According to our knowledge and after considering a wide range of CLV models with a clear segmentation proposal, all of them have a major drawback in common: CLV is not calculated through stochastic and disaggregated models that capture heterogeneity between customers as a first step of modelling. With this research we want to overcome this drawback through our first empirical stage, getting an accurate CLV measure for each customer (CLV_i, where i refers to each customer) and finally, as a second stage, applying regression trees technique in order to segment the customer base using CLV_i as an input variable. Therefore, we select the ideas related to option 2 to segment our sample of customers.

5.2.2. Value-based segmentation for this research

Once we have calculated each CLV_i, where i refers to each customer, we get a customer classification based on that value and certain socio-demographic variables (age_i, $gender_i$ and $income_i$) using data mining techniques. In particular, a decision tree is used (e.g., Hwang *et al*., 2004; Kim *et al*., 2006; Paauwe *et al*., 2007) to perform this second research stage. More specifically, we use regression trees.

As far as segmentation is concerned, there are two key concepts: a segmentation basis and a segmentation method. According to the basis of segmentation, it is defined as the identifying of "*a set of variables or characteristics used to assign potential customers to homogeneous groups*" (Wedel and Kamakura, 2000 p. 7). Wedel and Kamakura identify several segmentation schemes classified as we have shown in Table 9. In this scheme, 'observable' segmentation criteria are those that can be directly and objectively measured, whereas the 'unobservable' category implies the need for inference from available data.

Table 9. Bases of segmentation schemes (*)

	General	**Product specific**
Observable	Geographic, demographic, and socio-economic variables	User status; frequency of use; store loyalty
Unobservable	Psychographics; values; personality; life-style	Benefits; perceptions; preferences; intentions

(*) Source: adapted from Wedel and Kamakura (2000)

The bases of segmentation schemes previously shown will clearly differ in terms of their effectiveness but in practice, the eventual choice will depend on the purpose of the market study, the nature of the market, and the choice of segmentation methods. For this research, we have mixed socio-demographic variables with individual CLV's (that are defined from product specific variables). Therefore, our segmentation scheme uses observable variables, combining general and product specific information.

Additionally, Beane and Ennis (1987) summarised several segmentation methods and modelling techniques prevalent two decades ago, in particular: automatic interaction detection and its multivariate variant; canonical analysis; factor analysis; cluster analysis; regression analysis; discriminant analysis; multidimensional scaling; conjoint analysis and componential segmentation. Moreover, Wedel and Kamakura (2000) reviewed more recently developed methods, such as log-linear and mixture regression models, and classified them into the next Table 10.

Table 10. Classification of segmentation methods

	A priori/*Ex ante*	**Post hoc/*Ex poste***
Descriptive	Log-linear models	Non-overlapping (K-means)
	Contingency tables	Overlapping
		Fuzzy techniques
Predictive	Regression	Automatic interaction detection
	Discriminant analysis	Mixture regression models

(*) Source: adapted from Wedel and Kamakura (2000)

Regarding the classification previously shown, firstly, an a priori or an *ex ante* segmentation technique implies that "*the type and number of segments are determined in advance by the researcher*"; on the contrary, a post hoc or an *ex poste* segmentation technique implies that "*the type and number of segments are determined on the basis of the results of data analyses*" (Wedel and Kamakura, 2000 p. 17). Secondly, a descriptive segmentation technique consists of applying statistical methods that are descriptive, analysing associations among a set of variables but making no distinction between dependent and independent variables, whereas for a predictive segmentation technique one set consists of dependent variables to be explained/predicted by the other set of independent variables (Wedel and Kamakura, 2000 p. 17). The choice among these methods will depend on the objectives of the research.

On the other hand, within the context of CRM, data mining can be seen as a business driven process aimed at the discovery and consistent use of profitable knowledge from organizational data (Ling and Yen, 2001), for example, data mining can increase the response rates of a marketing campaign by segmenting customers into groups with different characteristics. Therefore, the customer classification requirements of a CRM can be supported by different data mining models, such as decision trees. This customer classification aims at building a model to explain or even predict future customer behaviour through classifying database records into a number of predefined classes based on certain criteria (Ngai *et al*., 2009).

In our case, we perform an *ex poste* segmentation to classify customers using data mining techniques. We define CLV_i as the dependent variable and this dependent variable is explained by a set of independent variables (all of them are observable variables), age_i, $gender_i$and $income_i$. The objective is to get profiles of customers according to the value of CLV. Various classification techniques are available, all of them aiming to explain/predict the response behaviour of individuals as accurately as possible. In particular, tree models are computationally intensive methods that are used in situations where there are many explanatory variables and the researcher needs guidance about which of them to include in the model (Crawley, 2007 p.685). The great virtues of these models are:

- Tree models are very simple.
- Tree models are excellent for initial data inspection.
- Tree models give a very clear picture of the structure of the data.
- Tree models provide a highly intuitive insight into the kinds of interactions between variables.

This group of techniques, called decision trees or simply tree models, is composed by automatic interaction detection (AID), chi-squared automatic interaction detection (CHAID) and classification and regression trees (CART). All of these algorithms divide a data set in exclusive and exhaustive segments that differ with respect to the response or dependent variable. The results of these three algorithms are a decision tree structure with a split in each node. The leaves, final or terminal nodes (because they cannot be split any further) are defined as combinations of the used independent variables. Furthermore, the leaves contain different distributions of the dependent variable (Van Diepen and Franses, 2006).

On one hand, we have established a comparison between AID and CHAID. Both methods belong to a family of methods known as Automatic Interaction detection (AID). As its name suggests, the AID allows for the detection of interactions between variables. Thus, the segmentation is based on the interactions. The AID requires that predictors are categorical, i.e., either discrete or discretised (if originally continuous) (Bijak and Thomas, 2012). In particular, AID operates on an interval scaled dependent variable and maximizes the between segment sum-of-squares at each bisection, whereas CHAID operates on a nominal dependent variable and maximizes the significance of a chi-squared statistic at each partition, which is not necessary a bisection. CHAID will merge those categories of a predictor that are homogeneous with respect to the dependent variable, but will maintain all categories of a predictor that are heterogeneous. Since more than two categories of a predictor may differ significantly, the results of the CHAID merging process need not to be a dichotomy. Furthermore, CHAID draws on the theory of combinatorial statistic as it applies to the use of Bonferroni multipliers in adjusting probability levels for multiple hypothesis tests on the same data (Hawkins and Kass, 1982), that is, the Bonferroni correction adjusts the test significance level for many tests that are performed at the same time. By using the Bonferroni adjustment, CHAID makes up for the fact that a number of original categories can be merged into a smaller number of combined categories in various ways. The adjustment nullifies the bias towards predictors with more categories. Finally, as opposed to AID, CHAID will only split the data on a specific predictor if this leads to a significant difference in distribution of the dependent variable. In this way, the sampling variability in the data is taken into consideration. To sum up, CHAID partitions the data into mutually exclusive, exhaustive, subsets that best describe the dependent variable (Van Diepen and Franses, 2006).

On the other hand, although the theoretical foundations between CART and CHAID are quite different, the two techniques perform very similar, in that they produce similar results as far as increases in response are concerned. Overall, CART is preferred when there are many continuous variables and CHAID when there are many categorical variables (Van Diepen and Franses, 2006). Therefore, we select CART or regression trees in order to perform our proposed segmentation (see Appendix B for more details about the background of the regression tree algorithm applied by R software). In this case the response variable is a continuous measure (i.e., CLV_i, where i refers to each customer), but the explanatory variables can be any mix of continuous and categorical variables (i.e., continuous variables such as age_i and $income_i$ and categorical ones such as $gender_i$).

5.3. Final empirical stage: *COMPUTING CE FOR THE ENTIRE CUSTOMER BASE*

Finally, we calculate CE as a proxy for the overall assessment of the company. An important empirical illustration which firstly calculates CLV, secondly carries out a segmentation (but simpler than ours) and compute CE at the end of the process can be found in Kumar and Shah (2009).

As we have noted in Chapter 2, we define CE in the same line of Zhang *et al*. (2010): ***CE is formed by the CLV's of all customers in the database***, which has been found to be a ***good proxy measure of the equity-market valuation of the firm*** (Gupta *et al*., 2004). In general, we refer to a ***dynamic CE*** (which is defined as the discounted sum of both current and future cohorts' CE).

For this research CE is specified as follows:

$$CE = mean(CLV_i) * POP = \sum_{i=1}^{I} CLV_i\,, if\ POP = sample$$

Where:

mean (CLV_i) = average lifetime value for firm customers (i) across the sample and,

POP = total number of customers in the sample.

5.4. Concluding remarks

In this dissertation we present a model for the assessment of customers, which we have developed in cooperation with a leading Spanish retail bank. This model is based on a combination of a hierarchical Bayesian model to estimate CLV and regression tree analysis to examine several important research questions: **Which drivers of CLV have more potential to predict components of CLV?**, **what is the long-term value of each customer (CLV_i)?**, **what is the value of the customer base (CE)?** and **which groups of customers are more (less) valuable?**

An in-depth understanding of these questions is of interest and importance to both managers and researchers, since the results can be used as input to marketing decisions, which for example, can contribute to acquire economic returns from customers as an important asset of the company. To serve these interests, a model in Figure 3 (see Chapter 4) is built to develop an integrated framework to estimate individuals CLV's as a basis for segmentation. The model has been validated using a sample of 1.357 customers with 32.568 datasets from a leading Spanish retail bank (for more details see the next Chapter).

Chapter 6. EMPIRICAL STUDY

In this section we present the application of our methodology to a real-life setting: a Spanish financial multi-service retailer. This section is organised as follows. Firstly, we justify the choice of this context and we provide further details and information about the retail financial services setting. After that, data available and variables measured are described. Finally, we provide the results of the empirical application of the model.

6.1. Research context

We illustrate our new procedure to assess and segment customers with data from a Spanish financial multi-service retailer. This financial services firm has a wide range of products and services, including saving, insurance and investment products. We have selected this context because nowadays the financial markets have become more competitive due to several reasons, for instance the mature nature of the sector (Prinzie and Van den Poel, 2006), deregulation, new competitors (Ritter, 1993) and the intensive European financial integration (Dawes and Swailes, 1999). A new environment has been created which allows customers to choose among a wide range of options to meet their financial needs (Colgate and Danaher, 2000). As a consequence, the number of new entrants in the retail banking sector has risen, coming from industries as diverse as insurance and automobile production (Haenlein *et al.*, 2007). This fact results in the commoditization of basic banking products, such as deposits, mortgages and credit extensions, in diminishing profit margins and the blurring distinctions between banks, insurers and brokerage firms (i.e., universal banking).

Moreover, the situation of the sector was heavily damaged by the so-called subprime crisis (BusinessWorld, 2007; The Washington Post, 2007), which has generated a more critical environment for banks. Under this intensive competitive pressure, companies realise the importance of retaining their current customers. In particular, the retail banking sector, which is predominantly service based, derives higher profits from the creation and retention of long-term relationships with customers. The substantive relevance of CLV modelling comes from the fact that an increase in retention rate of the best customers of just one percentage point may result in substantial profit increases (Van den Poel and Larivière, 2004). Hence, nowadays a small number of large institutions offering a wider set of services dominate the financial-services industry. These CLV developments stimulated banks to implement Customer Relationship Management

(CRM), because banks should take special care with those customers that 'allow' their existence. The nature of the banking industry is risky by itself due to uncertainties about the ability of customers in being able to pay back for example their loans (or not), so it is crucial for a financial service provider to recognise which relationships are hazardous and should be avoided.

In order to get a better understanding of the importance of CLV in this context, we refer here to "*The Spanish Banking Study 2012*", which was developed by IBM (with the cooperation of most Spanish banks) during the last quarter of 2009, in order to set priorities for the sector (IBM, 2010). Among its conclusions we highlight: the development of a better customer management, focusing on strengthening relationships with existing customers through excellence in service quality; the development of a customer-centric banking; using the trend towards concentration and restructuring of the sector to acquire new customers; considering the processes of collecting and managing customer information as a source of competitive advantage; and to improve the use of technology to manage the customer experience.

Additionally, another reason that justifies our choice is that among CLV research we have found a common suggestion for further research, i.e., apply the different models in other types of business relationships, especially in the financial services context (Lewis, 2006). There is also a necessity of models that cover customer's relationships with a portfolio of the products of the company (Rust and Chung, 2006), or in other words, models that deal with different product categories (Jain and Singh, 2002). The purpose of the previously mentioned suggestion is, for example, to examine the effects of marketing dynamics on CLV and CE, e.g., cross-selling between a multi-product brands or products of the firm (Aravindakshan *et al*., 2004). This task constitutes a challenge for this research. This is because despite the apparent theoretical simplicity of the CLV concept, it is fraught with difficulty when applied in practice, in particular in a banking context. In such a context, purchase behaviour is rather complex because customers can purchase more than one service (there are a large number of heterogeneous services at their disposal). These purchases are often not independent from each other. It is difficult to assign an amount of profits or margin to each transaction because of the complicated financial conditions in this sector and, additionally, there are different types of transactions and channels available to customers with different fees, etc.

The retail banking sector is also a context that is particularly suitable for research on CLV due to the prevalence of sophisticated CRM systems (Krasnikov *et al*., 2009), the comparatively long duration of many relationships (Leverin and Liljander, 2006) and the fact that, due to customer acquisition costs, it can take several years for customers to break even and become profitable (Reichheld, 1996). Therefore, the banking sector is a suitable context to examine our proposed model.

6.2. Data set

As we have noted previously, a Spanish financial services retailer provides the dataset for the purposes of our analysis. Data pre-processing was required to ensure data field consistency because, obviously, not all the data are related to the chosen purposes. We have extracted only those data considered being useful for the analysis. Unnecessary data fields have been removed from the database and missing data have been imputed by *multiple imputation* using IBM SPSS Statistics version 20.

6.2.1. Treatment of missing data

According to the nature of missing data, the literature (e.g., Rubin, 1987; Schafer and Graham, 2002) lists three possibilities of 'missing-ness'. For an easier understanding, we describe these statistical terms intuitively (for a precise statistical definition, see Rubin (1987)). Consider a data set with a number of variables ($X_1, ..., X_n, Y$). Assume that Y displays some degree of missing data:

- The missing data for Y are said to be completely at random (MCAR) if the probability that Y is missing is unrelated to the values of Y itself and is unrelated to any of the other variables in the data set.

- Missing data are said to be missing at random (MAR) if the probability of missing data in Y is unrelated to the values of Y, after the other variables in the data set are controlled for. The missing data may be related to the values of $X_1, ..., X_n$.

- The missing data are said to be missing not at random (MNAR) if the probability that Y is missing depends on the value of Y.

The assumptions about the types of 'missing-ness' are important because they determine what can be expected from the performance of the statistical technique used to handle missing data. Therefore, in order to determine if our missing data are MCAR, the χ^2 test developed by Roderick J. A. Little (Little, 1988) was applied using the missing data module in SPSS version 20. The goal of this first analysis is to test the following null hypothesis: the data are completely missing at random. Our results are significant (χ^2 = 4408,46; d.f. = 506, p < 0,05), therefore we have rejected the null hypothesis, or in other words, our data are not missing completely at random. Secondly, MAR is an un-testable assumption (unless we have knowledge of the missing values themselves). In order to find possible dependence in missing patterns, we have inspected these patterns to determine approximately if our missing data are MAR.

According to the analysis of the missing patterns, missing data are presented in the following variables (where i is the customer index and the total number of observations are 32.568 from 1.357 customers):

- ***Income***, with 1.974 missing observations (6,06% over the total number of observations).
- ***Average monthly assets***, with 50 missing observations (0,15% over the total number of observations).
- ***Average monthly liabilities***, with 50 monthly observations (0,15% over the total number of observations).

In Figure 8, we show the missing data patterns of the variables in the analysis. This analysis was performed also using the missing data module in SPSS version 20. A missing pattern defines the way by which missing values are generated and helps us to choose an imputation method. If the database is interpreted as a matrix, to define a missing pattern the columns are the observation units and the rows represent the variables of interest. Each column represents a different missing pattern, which defines groups of variables with missing values (they are highlighted in black). Variables and missing patterns are arranged in order in this matrix to reveal trends of monotony and randomness, in particular from those variables with more missing observations (at the top-left part) or patterns with more missing observations (the first column on the left) to those variables with less number of missing values (at the bottom-right part of the matrix) or patterns with less number of missing values (columns on the right part of the matrix). With the help of Figure 8, we can conclude that our missing values, instead of being monotonous, they present a random pattern (we can find missing data in any cell). Additionally, variables with missing values present some level of dependence between them. Therefore, we are going to assume that our missing data are missing at random (MAR).

Figure 8. Missing data patterns of the variables in the analysis (missing data in black)

	Missing patterns							
Variables	8	7	6	5	4	3	2	1
INCOME								
AVMONTHLIABILITIES								
AVMONTHASSETS								
VARIABLES WITHOUT MISSING								

Missing data, also referred to as missing values, is a bigger problem than is not often recognised or is not given enough attention and importance (Peugh and Enders, 2004). Many studies have demonstrated that missing data can lead to biases in statistical results. In particular, Collins *et al.* (2001) state: "*When researchers are confronted with missing data, they run an increased risk of reaching incorrect conclusions. Missing data may bias parameter estimates, inflate Type I and Type II error rates, and degrade performance confidence intervals. Furthermore, because a loss of data is nearly always accompanied by a loss of information, missing values may dramatically reduce statistical power*". Therefore, when a researcher estimates a model with an incorrect treatment of missing data, the results can be biased. Furthermore, because the research findings may be strongly linked to specific marketing decisions, such biases can lead to substantially incorrect or suboptimal marketing decisions (Grover and Vriens, 2006). Thus, treatment of missing data is an important task that has to be performed by the researcher previously to implementing the model. The knowledge about how to impute missing data helps researchers make the right recommendations and it potentially helps them better leverage existing information.

6.2.1.1. Simple methods

The simplest ways to deal with missing data consist of (i) no-treatment option and (ii) single imputation methods. On one hand, (i) no treatment option includes ***'complete case analysis'* or *'listwise deletion'***. This first option implies to use only those individuals who have no missing data on the variables that we want to analyse. The ***'available case method'* or *'pairwise deletion'*** is another way to not deal with missing data. It uses only the data that are available. In general, pairwise deletion is less preferred than the listwise deletion (Allison, 2001). Pairwise deletion is usually used in conjunction with a correlation matrix. Each correlation is estimated based on the cases having data for both variables. The issue with pairwise deletion is that different correlations (and variance estimates) are based on different subsets of cases. Because of this, it is possible that parameter estimates based on pairwise deletion will be biased. Additionally, because different correlations are based on different subsets of cases, there is no guarantee that the matrix will be positive definite. Nonpositive definite matrices cannot be used for most multivariate statistical analyses. Another concern with pairwise deletion is that there is no basis for estimating standard errors (Graham, 2009). Because of all these problems, some authors cannot recommend pairwise deletion as a general solution (e.g., Graham, 2009). In general, no treatment approaches can become problematic for a number of reasons: (1) when the missing data are not missing completely at random (MCAR), the estimated statistics, for example regression coefficients, will be biased and inefficient (Ramaswamy *et al*., 2001); and (2) it will lead to a large amount of data being discarded.

On the other hand, among (ii) single imputation methods we can find ***'mean substitution method'***, which substitutes the computed available cases mean for all cases without missing data for the variable under consideration. It is simple to implement but it is generally not recommended because it has the following disadvantages: (1) it distorts the underlying distribution of the data, making the distribution more peaked around the mean and reducing the variance of the variable; and (2) it does not take into account that the imputed data have more uncertainty than does a complete data set (Grover and Vriens, 2006).

Another single imputation method is ***'single regression substitution'***. To apply this method, the researcher attempts to make use of the structure in the data, therefore it is conceptually better than the previously discussed methods. In particular, it performs the following steps:

- For the variable that has missing data, an available case regression is performed with that variable as the dependent variable. Other variables in the data set are predictors.
- The missing values then are replaced by the corresponding values that the estimated regression equation predicted.
- This procedure is repeated for all variables that are incomplete as a result of missing data.

In general, regression substitution is better than mean substitution because it can yield consistent estimates, that is, the estimates will be unbiased with large sample sizes if the missing data are MAR. However, it assumes no residual error around the regression line, and therefore it underestimates the variability of the estimates.

6.2.1.2. Advanced methods

The relatively simple methods discussed previously all have undesired statistical properties. Fortunately, to apply a better treatment of missing data, researchers can choose advanced approaches that have better statistical properties for dealing with this real problem.

Among these advanced approaches, the ***'expectation maximization (EM) method'*** is an interesting option to consider, especially if your data are not MCAR (IBM, 2011 p. 3). The EM algorithm consists of an iterative procedure that generates estimated values for missing data by using expectation (E-step) and maximization (M-step) algorithms (Hedderley and Wakeling, 1995), and a covariance matrix and mean vector are subsequently estimated. Following the explanation about EM algorithm given by Enders (2003), the process starts with an initial estimate of the covariance matrix, which is used to construct a series of regression equations in the first E step. For example, consider a simple scenario with four variables, x_1 through x_4. For cases with missing values on x_1, the covariance matrix is used to generate the regression of x_1 on x_2 through x_4. Similarly, for cases missing both x_1 and x_2, the regression of x_1 on x_3 and x_4 and the regression of x_2 on x_3 and x_4 would be obtained. The predicted values generated from these regression equations essentially serve as estimates of the missing data points using all variables in the data (with and without missing values). After all missing values have been imputed in the E step, the resulting data matrix is used to obtain updated estimates of the covariance matrix and mean vector in the M step. The updated covariance matrix computed at the M step is, in turn, used to construct a new set of regression equations in the next E step (Enders, 2003). The algorithm iterates through these steps until the difference between covariance matrices in successive M steps falls below some convergence criterion (Enders, 2003) or in other words, the process iterates until changes in expected values from iteration to iteration become negligible (Hedderley and Wakeling, 1995).

The EM approach is considered superior to listwise, pairwise and mean substitution methods and is assumed to produce unbiased parameter estimates for MCAR and less biased parameter estimates for MAR and nonignorable missing data (Acock, 1997; Musil *et al*., 2002). This method has limitations in that the parameters (means and covariances) generated by the EM procedure are reliable and correct, but standard errors are lower and therefore some test statistics (e.g., t tests) may be inaccurate (Allison, 2002; Musil *et al*., 2002). More accurate methods, for instance, multiple imputation (which we review in next paragraphs), are computationally more intensive and involves approximations, therefore it cannot deal with a small amount of missing values because the method needs an enough amount of missing values in order to run the simulation (IBM, 2011).

Another advanced approach is the ***'stochastic regression imputation'***. The single imputation version of this approach implies an improvement over the regression substitution method that was mentioned above. This single imputation method replaces each missing value in the data set by a predicted value from a regression analysis, based on available cases and a random residual term. However, a single imputation method underestimates the variability in the data. Therefore, if the researcher employs multiple regression, each missing value is replaced by a set of $m > 1$ plausible values to generate m complete data sets. The stochastic regression imputation method can become quite complicated when there are multiple patterns of missing data, because different regression equations must be constructed for each unique pattern (Peugh and Enders, 2004 p. 529). This may be the reason this method is not widely adopted.

Unfortunately, the two previously mentioned advanced methods (i.e., expectation maximization and multiple regression) can impute missing values in case of quantitative variables only. However, our missing values are presented in quantitative variables but with constraints, such as being positive (e.g., income, average monthly assets and average monthly liabilities). For this reason, we have chosen the ***'multiple imputation method'*** to impute missing values for the three variables previously mentioned (i.e., ***income, average monthly assets*** and ***average monthly liabilities***). This method can deal with any kind of data (Allison, 2011), including those values that are not missing at random (IBM, 2011 p. 49).

In particular, multiple imputation appears to be one of the most attractive methods for general-purpose handling of missing data. The basic ideas, proposed by Rubin (1987) are simple:

- Impute missing values using appropriate model that incorporates random variation (e.g., linear regression or logistic regression).
- Use Monte Carlo simulations to do this process M times (usually three to five times), producing M 'complete' data sets.
- Perform the desired analysis on each data set using standard complete-data methods.
- Average the values of the parameter estimates across the M samples to produce a single-point estimate.

- Calculate the standard errors by (a) averaging the squared standard errors of the *M* estimates, (b) calculating the variance of the *M* parameter estimates across samples, and (c) combine the two quantities using a formula.

Multiple imputation has several desirable features (Landerman *et al*., 1997; Little and Rubin, 1989):

- Introducing appropriate random error into the imputation process makes it possible to get approximately unbiased estimates of all parameters. No deterministic imputations can do this in general settings.

- Repeated imputation allows one to get good estimates of the standard errors. Single imputation methods do not allow for the additional error introduced by imputation (without specialised software of very limited generality).

- Multiple imputation can be used with any kind of data and any kind of analysis without specialised software.

Different multiple imputation (MI) algorithms have been proposed (Allison, 2001). We are focused on the automatic method that SPSS version 20 offers to us. Therefore, to perform this analysis we have used missing value module in SPSS running five simulations performed in sequence (with one hundred iterations each one). At the end of this process, the values are averaged together to take into account the variance of the missing values and to get a single value to impute in the empty cells. As we have noted at the beginning of this paragraph, we have selected the automatic method to get missing values. This way to proceed implies that SPSS applies linear regression to estimate missing values for quantitative variables (in our case, income, average monthly assets and average monthly liabilities) and logistic regression to estimate missing values for categorical variables. We have also fixed some restrictions to the imputation of missing values, such as income, average monthly assets and average monthly liabilities that cannot have a negative value (for categorical variables it is not necessary to fix any restrictions because they are automatically dummy coded).

6.2.2. Description of the data set

Once we have solved the problem of missing data, the database contains 24 months of behavioural data for 1.357 customers with 32.568 datasets. The time period of observation considered in this database begins on December 2010 and ends on November 2012. Theoretically, CLV models should estimate the value of a customer over the entire customer's lifetime, but in many firms, including our collaborating firm, two years of data are considered a good time horizon over which the current business environment would not change substantially. All customers started their relationships with the bank during this period, that is, we only use those customers who started purchasing from the multiservice retailer since December 2010, not before. This decision imply to ensure the data were not left-censored (i.e., to ensure the customers in our database really started their relationship with the company at the time of our first observation) (Baesens *et al*., 2004). In summary, from the total base of customers we have selected those ones who made their first transaction in December 2010 for the purpose of the CLV prediction. The choice of the sample and the subsequent two years as the observation period available has certain limitations. Such limitations have an impact on age of customers (generating bias towards young families), but also on length of the relationship and the opportunity for cross-selling.

For each month and each customer the following information is observed: type, frequency and total quantity of product purchased, length of the relationship, average monthly assets, average monthly liabilities, adoption of online banking and profitability. Socio-demographic information of each customer is also observed: age (as continuous variable), gender (as binary variable where "1" = male, "2" = female) and income (as continuous variable).

For the independent variables, we include past behavioural data. We have incorporated as many predictors as possible in order to determine which type of information is the most important in predicting CLV (for further details see Chapter 4). Table 11 gives an overview of the different variables that are used in the analysis. All these variables have proven to be good predictors of lifetime value in previous studies and additionally they are chosen because they are of special interest for the company.

Table 11.1. Overview of the variables in the analysis (*)

LIST OF VARIABLES		**Quantitative**	**Qualitative**	**Time varying**	**Non-time varying (**)**
Components of CLV (dependent variables)					
Product Ownership: O_01_{it}… O_18_{it}			binary variable	X	
Product usage: U_01_{it} … U_18_{it}		X		X	
Contribution margin: CM_{it}		X		X	
Drivers of CLV (independent variables)					
Frequency variables: one-period lagged variables of product ownership (product ownership at (t-1)): O_{ijt-1}	Stock capital: O_01_{it-1}		binary variable	X	
	Credit Card: O_02_{it-1}		binary variable	X	
	Debit Card: O_03_{it-1}		binary variable	X	
	Saving Insurance: O_04_{it-1}		binary variable	X	
	Home Insurance: O_05_{it-1}		binary variable	X	
	Not Linked Life Insurance: O_06_{it-1}		binary variable	X	
	Linked Life Insurance: O_07_{it-1}		binary variable	X	
	Other Insurances: O_08_{it-1}		binary variable	X	
	Account: O_09_{it-1}		binary variable	X	
	Home Loan: O_10_{it-1}		binary variable	X	
	Deposit: O_11_{it-1}		binary variable	X	
	Investment Fund: O_12_{it-1}		binary variable	X	
	Pension Plan: O_13_{it-1}		binary variable	X	
	Securities: O_14_{it-1}		binary variable	X	
	Consumer Loan: O_15_{it-1}		binary variable	X	
	Micro-consumer Loan: O_16_{it-1}		binary variable	X	
	Mortgage: O_17_{it-1}		binary variable	X	
	Credit: O_18_{it-1}		binary variable	X	

(*) Where *i* refers to customers and *t* refers to periods of time.

(**) Variables that are non-time varying are measured at the end of the observation period (such is the case of length of the relationship and age) or are constant variables (such is the case of gender).

Table 11.2. Overview of the variables in the analysis (continued) (*)

LIST OF VARIABLES		**Quantitative**	**Qualitative**	**Time varying**	**Non-time varying (**)**
Drivers of CLV (independent variables) (continued)					
One-period lagged variables of product usage (product usage at (t-1)): U_{ijt-1}	Stock capital: U_01_{it-1}	X		X	
	Credit Card: U_02_{it-1}	X		X	
	Debit Card: U_03_{it-1}	X		X	
	Saving Insurance: U_04_{it-1}	X		X	
	Home Insurance: U_05_{it-1}	X		X	
	Not Linked Life Insurance: U_06_{it-1}	X		X	
	Linked Life Insurance: U_07_{it-1}	X		X	
	Other Insurances: U_08_{it-1}	X		X	
	Account: U_09_{it-1}	X		X	
	Home Loan: U_10_{it-1}	X		X	
	Deposit: U_11_{it-1}	X		X	
	Investment Fund: U_12_{it-1}	X		X	
	Pension Plan: U_13_{it-1}	X		X	
	Securities: U_14_{it-1}	X		X	
	Consumer Loan: U_15_{it-1}	X		X	
	Micro-consumer Loan: U_16_{it-1}	X		X	
	Mortgage: U_17_{it-1}	X		X	
	Credit: U_18_{it-1}	X		X	

(*) Where *i* refers to customers and *t* refers to periods of time.

(**) Variables that are non-time varying are measured at the end of the observation period (such is the case of length of the relationship and age) or are constant variables (such is the case of gender).

Table 11.3. Overview of the variables in the analysis (continued) (*)

LIST OF VARIABLES		**Quantitative**	**Qualitative**	**Time varying**	**Non-time varying (**)**
Drivers of CLV (independent variables) (continued)					
One-period lagged variable of profitability (monetary value at (t-1)): CM_{it-1}		X		X	
Length of the relationship: L_i		X			X
Recency variables	Purchase Recency: PR_{it}	X		X	
	Cancellation Recency: CR_{it}	X		X	
Cross-buying : CB_{it}		X		X	
Intensity of product ownership variables	Average monthly assets: AMA_{it}	X		X	
	Average monthly liabilities: AML_{it}	X		X	
Total quantity of purchases (across all product categories): Q_{it}		X		X	
Adoption of online banking: OL_{it}			binary variable	X	
Value based segmentation variables					
Age: AGE_i		X			X
Gender: SEX_i			categorical		X
Income: INC_i		X			X

(*) Where *i* refers to customers and *t* refers to periods of time.

(**) Variables that are non-time varying are measured at the end of the observation period (such is the case of length of the relationship and age) or are constant variables (such is the case of gender).

6.3. Measuring variables

Before starting to define the different measures related to each construct, it is needed to define three indexes to control customers, periods of time (months) and bank products, they are described below:

(1) i = index for customers ($1 \leq i \leq I$, where I is the total sample size),

(2) t = index for periods of time or months ($1 \leq t \leq T$, where T is the end of the calibration or observation time frame),

(3) j = index for banking products ($1 \leq j \leq J$, where J is the total number of banking products).

6.3.1. Components of CLV

To measure ***product ownership at t_1… t_{12}***, we include binary variables for ownership of each product type j at each time period t by each customer i, which is represented by 18 variables called from O_01_{it}… O_18_{it}. This way to proceed is commonly used in the literature (e.g., Donkers *et al*., 2007; Verhoef and Donkers, 2001).

To measure ***product usage at t_1… t_{12}***, we include continuous variables that measure the number of products of each type j that each customer i purchases at each time period t. This information is represented by 18 variables called from U_01_{it}… U_18_{it} (e.g., Bolton and Lemon, 1999; Bolton *et al*., 2000).

The third driver of CLV is ***contribution margin at t_1… t_{12}***. This variable is measured as the difference between interest and fees charged to the customer minus the bank cost paid or the bank incomes earned (because of the bank invests the money of customers funds and other products) at the *Interbank Lending Market*. It is represented by one continuous variable called CM_{it} (e.g., Kumar and Shah, 2009).

About ***discount rate*** (d), Chang (2011) used 'market interest rates' as discount rate. In a financial service setting, the discount rate d depends on the general rate of interest and is normally proportional to the treasury bill or the interest that banks pay on savings accounts (Kumar, 2006). It can also vary across firms depending upon the cost of capital to the firm. In our case, d is equal to the euribor rate (in Table 12 euribor monthly rates are shown).

Table 12. Euribor, monthly rates

Months	2013	2012	2011	2010
January	0,575	**1,837**	**1,55**	1,232
February	0,594	**1,678**	**1,714**	1,225
March	0,545	**1,499**	**1,924**	1,215
April	0,528	**1,368**	**2,086**	1,225
May	0,484	**1,266**	**2,147**	1,249
June	0,507	**1,219**	**2,144**	1,281
July	0,525	**1,061**	**2,183**	1,373
August	0,542	**0,877**	**2,097**	1,421
September	0,549	**0,74**	**2,067**	1,42
October		**0,65**	**2,11**	1,495
November		**0,588**	**2,044**	1,541
December		0,549	**2,004**	**1,526**
TOTAL	2,242	13,332	24,07	16,203

Source: www.euribor.com.es/valor-euribor

6.3.2. Drivers of CLV

Purchase frequency variables measure the product choices (***product ownership***) the customer *i* has made from the bank in period *(t-1)* (Donkers *et al*., 2003; Donkers *et al*., 2007). This is a one-period lagged variable that is operationalised as time-varying. We include binary variables for ownership of each product type j (where 0 means no product ownership and 1 product ownership). They are a total of 18 time-varying binary variables, called from O_01_{it-1} *to* O_18_{it-1} (Donkers *et al*., 2007).

Product usage variables measure the total quantity of purchases of each product the customer *i* has made from the bank in period *(t-1)* (Bolton and Lemon, 1999; Venkatesan *et al*., 2007 p. 585). This is a one-period lagged variable that is operationalised as time-varying. We include continuous variables for usage of each product type j at *(t-1)* by each customer *i*, which is represented by 18 variables called from U_01_{it-1} to U_18_{it-1}. In order to reflect the latent nature of this predictor variable without losing the detailed information provided by their respective items, we have calculated an additional predictor variable (detailed as Q_{it}) by summing up all or sub-sets of their formative indicators. In particular, it measures ***total quantity of purchases*** by customer *i* at period *t* (Venkatesan and Kumar, 2004) across all product categories. This approach is in line with the basic philosophy behind this type of measurement (see Jarvis *et al*. (2003) for more details).

The one-period lagged variable of ***contribution margin*** (CM_{it-1}) earned by the bank from each customer i in period *(t-1)* is measured as the difference between interest and fees charged to the customer minus the bank cost paid or the bank incomes earned (because of the bank invests the money of customers funds and other products) at the *Interbank Lending Market*. This is a lagged variable that is operationalised as time-varying (Kumar and Shah, 2009).

Length of the relationship *(L_i)* is defined as the time between the entry of the individual i as a customer of the company until the end of the observation period (Glady *et al*., 2009a). To measure length of the relationship we use a continuous variable (in months), which indicates the length of the duration of the relationship for each customer *i* in the database (Reinartz and Kumar, 2003 p.88). Some customers in the sample end their relationship with the bank before the end of the observation period (they are a total of 179 customers).

For recency, we include two continuous variables, (1) ***purchase recency*** (PR_{it}), which measures the number of periods (months) since the last purchase across all product categories, and (2) ***cancellation recency*** (CR_{it}), which measures the number of periods (months) since the last cancellation across all product categories. Donkers *et al*. (2007) operationalised these two variables as binary variables, but we have followed the pure definition of recency (Pfeifer and Caraway, 2000) defining purchase recency as the difference between the last purchase and the (period of) time of analysis, and cancellation recency as the difference between the last cancellation and the (period of) time of analysis. Therefore, we use two continuous variables to measure purchase recency and cancellation recency instead of two binary variables (such is the case of Donkers *et al*., 2007).

Following the suggestions of Verhoef *et al*. (2001), we measure ***cross-buying*** as the difference in the number of products purchased/cancelled across all product categories from the focal supplier between t_{n+1} and t_n as reflected by the company database (CB_{it}). Note that the difference can also be negative (in case of cancellations). As suggested by Verhoef *et al*. (2001), we jointly consider these positive and negative scores because increasing the number of services and decreasing the number of services is the same decision process (Bolton and Lemon, 1999). We have included this variable because it offers a way to quantify the real number of products that each customer acquires (if cross-buying is a positive quantity) or cancels (if cross-buying is a negative quantity). Purchase recency and cancellation recency give us limited information about if a purchase or a cancellation occurred because they do not quantify the real number of products that are purchased or cancelled. For this reason we have measured cross-buying as another driver of CLV.

With regard to intensity of product ownership, we have taken into account two additional continuous and time-varying variables: (1) ***average monthly assets*** (AMA_{it}) (measured in euros), and (2) ***average monthly liabilities*** (AML_{it}) (measured in euros). The first one (AMA_{it}) is defined as the monthly average of outstanding balance on short and long-term credit accounts, loans, debt on current account and investment products (Prinzie and Van den Poel, 2006). The second one (AML_{it}) is defined as the monthly average of savings and investment products, credit on current account and monthly insurance fees (Prinzie and Van den Poel, 2006). Although these two variables are continuous dimensions, they are registered in the database at discrete moments in time, for example, at the end of the month for bank accounts and on a yearly basis for insurance products.

Regarding ***adoption of online banking*** or PC banking (as Hitt and Frei (2002) called it), we use binary variables for each customer i and each period t. In particular, OL_{it} takes on a value of 1 if customer i adopts the online channel in month t, and equal 0 for all non-adoption months (Campbell, 2006; Campbell and Frei, 2010).

6.3.3. Value based segmentation variables

We include age, gender and income as ***socio-demographic variables*** to be used for segmentation purposes. Prior research has also incorporated demographics in an attempt to explain customer value (Abe, 2009b; Kumar *et al*., 2006a; Reinartz and Kumar, 2000; Verhoef and Donkers, 2001). These socio-demographic variables are explained below.

- ***Age*** (AGE_i): as a continuous variable measured at the end of the time period of calculation.

- ***Gender*** (SEX_i): as a dichotomous variable (1 = male, 2 = female) measured at the beginning of the relationship between customer and bank by the own bank.

- ***Income*** (INC_i): as a continuous variable that gives information about the mean salary that each customer receives each month. For the bank it is easy to know the income if the client has set up his/her salary in a direct deposit or account (these are 50% of customers in their databases and data warehouses); otherwise the bank has to estimate income (in the other 50% of the cases). For this task the bank operates as follows. Firstly, [if the previously mentioned data are not available for the bank], the income is estimated from the information related to the incomes in current account/s. Secondly, [if the previously mentioned data are not available to the bank], income is estimated from the monthly amount for a mortgage. Thirdly, [if also the previously mentioned data are not available to the bank], the estimation of income is based on the monthly amount designated to pay a consumption loan. Fourthly, [if the previously mentioned data are not available to the bank], income is estimated from the movements of the current account/s. Finally, if any of the previously mentioned data are not available for the bank, a decision tree is developed to form groups according to similarities in socio-demographic characteristics among clients (without income information and with it). When groups of clients are obtained, an average income per group (from clients with income information) is assigned to the clients without income information in each group.

6.4. Results

6.4.1. Descriptive statistics

Descriptive statistics for all variables included in the model are shown in Tables 13. We show this Table in order to give the reader more information about the magnitude of the variables that we are working with.

Table 13.1. Descriptive statistics

Variables	Number of observations (N)	Minimum	Maximum	Mean	Standard deviation	Variance
Customer ID	32.568	1	1.357	n.a.	n.a.	n.a.
Age	32.568	2	100	39,57	19,51	380,75
Gender	32.568	1	2	n.a.	n.a.	n.a.
Income (with missing values)	30.594	0	642.939,15	15.836,58	24.273,20	589.212.489,14
Income (after multiple imputation)	32.568	0	643.298,15	15.268,46	23.259,99	541.027.281,90
Length of the relationship	32.568	1	24	22,49	4,64	21,5
Purchase recency	32.568	0	24	9,68	6,82	46,6
Cancellation recency	32.568	0	24	10,63	6,93	48,08
Cross-buying	32.568	-4	5	0,01	0,26	0,07
Average monthly assets (with missing values)	32.518	0	813.905,53	7.629,56	44.156,34	1.949.782.184,00
Average monthly assets (after multiple imputation)	32.568	0	813.905,53	7.194,87	42.761,98	1.828.587.261,00
Average monthly liabilities (with missing values)	32.518	0	446.715,52	8.148,02	24.453,21	597.959.387,50
Average monthly liabilities (after multiple imputation)	32.568	0	446.715,52	7.648,36	23.737,95	563.490.307,50
Adoption of online banking	32.568	0	1	n.a.	n.a.	n.a.
Total quantity of purchases	32.568	0	12	2,3	1,64	2,7

n.a. = not applicable

Table 13.2. Descriptive statistics (continued)

Variables	**Number of observations (N)**	**Minimum**	**Maximum**	**Mean**	**S.D.**	**Variance**
O_01	32.568	0	1	n.a.	n.a.	n.a.
O_02	32.568	0	1	n.a.	n.a.	n.a.
O_03	32.568	0	1	n.a.	n.a.	n.a.
O_04	32.568	0	1	n.a.	n.a.	n.a.
O_05	32.568	0	1	n.a.	n.a.	n.a.
O_06	32.568	0	1	n.a.	n.a.	n.a.
O_07	32.568	0	1	n.a.	n.a.	n.a.
O_08	32.568	0	1	n.a.	n.a.	n.a.
O_09	32.568	0	1	n.a.	n.a.	n.a.
O_10	32.568	0	1	n.a.	n.a.	n.a.
O_11	32.568	0	1	n.a.	n.a.	n.a.
O_12	32.568	0	1	n.a.	n.a.	n.a.
O_13	32.568	0	1	n.a.	n.a.	n.a.
O_14	32.568	0	1	n.a.	n.a.	n.a.
O_15	32.568	0	1	n.a.	n.a.	n.a.
O_16	32.568	0	1	n.a.	n.a.	n.a.
O_17	32.568	0	1	n.a.	n.a.	n.a.
O_18	32.568	0	1	n.a.	n.a.	n.a.
U_01	32.568	0	2	0,3	0,46	0,21
U_02	32.568	0	2	0,06	0,24	0,06
U_03	32.568	0	3	0,54	0,59	0,35
U_04	32.568	0	2	0,02	0,13	0,02
U_05	32.568	0	3	0,04	0,23	0,05
U_06	32.568	0	2	0,02	0,14	0,02
U_07	32.568	0	3	0,09	0,31	0,1
U_08	32.568	0	2	0,02	0,14	0,02
U_09	32.568	0	5	0,91	0,5	0,25
U_10	32.568	0	2	0	0,06	0
U_11	32.568	0	10	0,18	0,58	0,33
U_12	32.568	0	3	0	0,09	0,01
U_13	32.568	0	2	0,03	0,16	0,03
U_14	32.568	0	2	0,01	0,09	0,01
U_15	32.568	0	2	0,04	0,22	0,05
U_16	32.568	0	1	0,01	0,9	0,01
U_17	32.568	0	3	0,05	0,22	0,05
U_18	32.568	0	1	0	0,06	0
CM	32.568	-10.764,98	12.790,28	100,89	723,15	522.946,82

n.a. = not applicable

6.4.2. Convergence diagnostics

6.4.2.1. Product ownership model

The proposed hierarchical Bayes Bernoulli model for product ownership is estimated by means of a Markov Chain Monte Carlo (MCMC) technique based on Gibbs sampling to explore the posterior distribution of the parameters of interest using WinBUGS version 1.4.3, available via the WinBUGS project Webpage. This task is accomplished by entering the evidence provided by the observed results of the dependent variable (product ownership) and updating the prior distributions by means of the Bayes' Theorem using a MCMC-based procedure (for more details about the model code see Appendix A1).

4.000 iterations are chosen as a *burn-in period* after which another 40.000 iterations that are run to obtain the *posterior distribution* of the parameters of interest (i.e., $\theta_0, \dots, \theta_8$). In order to verify parameter convergence, we use different diagnostics, such as the diagnostic developed by Gelman and Rubin (Gelman and Rubin, 1992). The term convergence refers to whether the algorithm has reached its equilibrium (target) distribution (Ntzoufras, 2009 p. 41). If it is true, then the generated sample comes from the correct target distribution. Hence, monitoring the convergence of the algorithm is essential for producing results from the posterior distribution of interest. There are many ways to monitor convergence (Ntzoufras, 2009 p. 41, 120, 129). In particular, we monitor convergence using: (1) Monte Carlo error (MC error) and autocorrelations; (2) trace plots; (3) history plots; (4) quantiles plots; and (5) Gelman-Rubin diagnostic.

(1) The simplest way to check convergence is to monitor the **Monte Carlo error (MC error)** and the **autocorrelations** (see Tables 14 enclosed below and Figures 9 in Appendix C1). Firstly, MC errors measure the variation of the mean of the parameter of interest due to the simulation (Ntzoufras, 2009 p.130). Small values of MC error will indicate that the quantity of interest is calculated with precision. In particular, if the MC error value is low in comparison to its posterior summaries (especially its standard error), then the posterior density is estimated with accuracy (Ntzoufras, 2009 p.130). Increasing the number of iterations will decrease the MC error. As a rule of thumb, the simulation should be run until the MC error for each parameter of interest is less than about 5% of the sample standard deviation. Moreover, Ntzoufras (2009 p.120) have noted that we can assume convergence when the MC error is lower than the 1% of the corresponding posterior standard deviation, such is our case for this first group of models (i.e., product ownership). Secondly, we can further monitor convergence using autocorrelations (Ntzoufras,

2009 p.41). If autocorrelations are low, such is also our case, convergence is obtained in a relative low number of iterations. Monitoring autocorrelations is also very useful since low or high values indicate fast or low convergence, respectively (Ntzoufras, 2009 p.120).

Table 14.1. Estimation results for the product ownership (stock capital)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
theta0	0,60	997,60	7,90	-1.950	2,75	1.949	0,79%	
theta1	10,40	995,30	7,72	-1.979	16,87	1.955	0,78%	32.859,63
theta2	1,17	1.003,00	7,48	-1.941	-0,54	1.969	0,75%	3,22
theta3	-8,80	1.002,00	7,67	-1.979	-9,48	1.942	0,77%	0,00
theta4	0,26	1.002,00	7,80	-1.986	8,60	1.945	0,78%	1,30
theta5	3,70	996,60	8,52	-1.962	5,25	1.968	0,85%	40,49
theta6	1,91	1.001,00	8,15	-1.951	5,05	1.931	0,81%	6,75
theta7	-3,22	1.005,00	8,81	-1.976	-0,40	1.952	0,88%	0,04
theta8	13,32	998,40	7,91	-1.925	18,90	1.951	0,79%	609.259,77

Table 14.2. Estimation results for the product ownership (credit card)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
theta0	-9,15	989,10	7,87	-1.937	-12,09	1.966	0,80%	
theta1	4,31	994,20	7,90	-1.981	9,08	1.935	0,79%	74,66
theta2	0,32	1.000,00	7,63	-1.948	4,57	1.967	0,76%	1,38
theta3	-1,66	1.000,00	7,75	-1.941	-12,81	1.965	0,78%	0,19
theta4	-1,01	1.004,00	7,80	-1.984	3,70	1.950	0,78%	0,37
theta5	1,22	992,20	8,45	-1.950	6,55	1.937	0,85%	3,40
theta6	6,54	997,00	8,10	-1.935	4,72	1.965	0,81%	692,29
theta7	12,56	1.005,00	8,06	-1.966	14,82	1.961	0,80%	284.930,34
theta8	5,19	1.003,00	7,67	-1.939	4,91	1.956	0,76%	179,65

Table 14.3. Estimation results for the product ownership (debit card)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
theta0	-5,18	982,90	7,85	-1.913	-12,21	1.930	0,80%	
theta1	-1,23	999,30	7,90	-1.974	-8,15	1.948	0,79%	0,29
theta2	-1,92	1.000,00	7,61	-1.951	-10,91	1.976	0,76%	0,15
theta3	-5,73	999,30	7,88	-1.955	-7,66	1.953	0,79%	0,00
theta4	3,81	1.002,00	7,73	-1.981	9,35	1.958	0,77%	45,11
theta5	1,30	1.001,00	8,41	-1.958	7,06	1.954	0,84%	3,65
theta6	10,43	995,00	8,26	-1.944	14,01	1.967	0,83%	33.860,35
theta7	5,18	1.001,00	8,39	-1.955	8,69	1.955	0,84%	177,15
theta8	6,44	1.002,00	7,94	-1.955	4,62	1.957	0,79%	628,29

Table 14.4. Estimation results for the product ownership (saving insurance)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
theta0	-1,01	991,90	7,85	-1.932	-6,24	1.962	0,79%	
theta1	14,35	996,50	7,76	-1.988	23,07	1.951	0,78%	1.706.576,71
theta2	-3,47	1.001,00	7,40	-1.939	-10,58	1.969	0,74%	0,03
theta3	5,10	997,10	7,92	-1.942	-1,27	1.964	0,79%	164,35
theta4	2,03	1.001,00	7,86	-1.980	14,17	1.952	0,79%	7,64
theta5	2,45	1.002,00	8,50	-1.971	0,53	1.967	0,85%	11,54
theta6	5,28	1.001,00	8,17	-1.932	8,53	1.970	0,82%	196,17
theta7	10,68	1.005,00	8,29	-1.973	19,79	1.967	0,82%	43.477,55
theta8	-0,52	999,60	8,24	-1.934	2,22	1.950	0,82%	0,60

Table 14.5. Estimation results for the product ownership (home insurance)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
theta0	-15,21	987,10	7,89	-1.948	-15,92	1.936	0,80%	
theta1	7,13	995,40	7,90	-1.970	1,97	1.938	0,79%	1.253,88
theta2	-0,31	997,70	7,53	-1.933	-1,57	1.961	0,75%	0,73
theta3	-2,53	1.001,00	7,79	-1.967	-19,51	1.955	0,78%	0,08
theta4	8,05	1.008,00	7,99	-1.979	19,01	1.977	0,79%	3.118,17
theta5	-1,17	989,50	8,33	-1.935	1,21	1.957	0,84%	0,31
theta6	5,27	999,30	8,41	-1.959	9,07	1.947	0,84%	193,83
theta7	2,49	1.009,00	8,34	-1.982	15,77	1.952	0,83%	12,01
theta8	10,84	1.003,00	8,11	-1.929	15,63	1.958	0,81%	51.021,38

Table 14.6. Estimation results for the product ownership (not linked life insurance)

Parameter	**Mean**	**S.D.**	**MC error**	**2,50%**	**median**	**97,50%**	**MC error/S.D.**	**e^mean**
theta0	-8,30	993,90	8,23	-1.944	-8,44	1.932	0,83%	
theta1	3,18	992,50	7,66	-1.977	12,50	1.940	0,77%	24,00
theta2	1,26	1.007,00	7,47	-1.942	-0,43	1.987	0,74%	3,51
theta3	3,08	996,60	7,79	-1.948	1,68	1.961	0,78%	21,82
theta4	13,36	1.005,00	8,18	-1.971	22,18	1.954	0,81%	634.124,13
theta5	7,28	993,50	8,16	-1.953	12,15	1.955	0,82%	1.450,99
theta6	8,79	995,50	7,84	-1.926	7,53	1.952	0,79%	6.574,80
theta7	11,45	1.014,00	8,16	-1.980	10,29	1.971	0,80%	93.901,35
theta8	3,19	997,50	7,78	-1.938	4,45	1.964	0,78%	24,19

Table 14.7. Estimation results for the product ownership (linked life insurance)

Parameter	**Mean**	**S.D.**	**MC error**	**2,50%**	**median**	**97,50%**	**MC error/S.D.**	**e^mean**
theta0	-8,73	986,40	7,58	-1.932	-9,38	1.947	0,77%	
theta1	4,17	992,90	7,74	-1.980	-1,79	1.934	0,78%	64,39
theta2	4,91	1.001,00	7,79	-1.932	7,65	1.970	0,78%	136,18
theta3	-0,12	1.001,00	7,86	-1.941	-15,27	1.978	0,79%	0,89
theta4	-0,54	1.007,00	7,91	-1.985	9,73	1.945	0,79%	0,58
theta5	1,78	1.002,00	8,18	-1.941	1,66	1.948	0,82%	5,90
theta6	3,83	998,90	8,12	-1.935	2,98	1.969	0,81%	46,11
theta7	13,80	1.003,00	8,10	-1.960	12,29	1.958	0,81%	984.609,11
theta8	1,74	1.012,00	8,01	-1.969	19,51	1.949	0,79%	5,71

Table 14.8. Estimation results for the product ownership (other insurances)

Parameter	**Mean**	**S.D.**	**MC error**	**2,50%**	**median**	**97,50%**	**MC error/S.D.**	**e^mean**
theta0	-7,64	999,20	7,91	-1.952	2,22	1.955	0,79%	
theta1	12,37	997,40	7,93	-1.979	16,24	1.946	0,80%	235.625,75
theta2	-9,39	1.007,00	7,92	-1.955	-13,44	1.967	0,79%	0,00
theta3	-9,97	996,40	7,81	-1.964	-14,44	1.941	0,78%	0,00
theta4	-1,38	1.007,00	7,72	-1.991	-1,65	1.948	0,77%	0,25
theta5	-7,25	999,90	8,20	-1.973	-4,63	1.962	0,82%	0,00
theta6	4,20	1.007,00	8,00	-1.960	9,52	1.969	0,79%	66,55
theta7	2,79	1.010,00	8,12	-1.962	1,72	1.965	0,80%	16,20
theta8	10,84	1.006,00	8,16	-1.935	17,23	1.957	0,81%	51.021,38

Table 14.9. Estimation results for the product ownership (account)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
theta0	-10,54	994,90	8,08	-1.957	-5,09	1.966	0,81%	
theta1	5,94	1.000,00	7,77	-1.974	6,64	1.944	0,78%	381,46
theta2	1,74	1.001,00	7,33	-1.928	1,25	1.981	0,73%	5,71
theta3	-4,27	998,10	7,93	-1.945	-8,96	1.949	0,79%	0,01
theta4	8,29	1.010,00	7,73	-1.973	15,73	1.959	0,77%	3.963,96
theta5	-6,31	994,60	8,15	-1.984	1,12	1.934	0,82%	0,00
theta6	8,92	1.001,00	8,18	-1.936	7,83	1.970	0,82%	7.465,14
theta7	13,64	1.005,00	8,65	-1.963	21,87	1.979	0,86%	839.028,54
theta8	6,57	1.005,00	8,28	-1.948	18,18	1.974	0,82%	712,66

Table 14.10. Estimation results for the product ownership (home loan)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
theta0	-1,05	993,60	8,04	-1.925	2,96	1.939	0,81%	
theta1	6,46	1.003,00	7,69	-1.986	11,85	1.936	0,77%	640,34
theta2	-2,25	1.003,00	7,95	-1.958	5,25	1.981	0,79%	0,11
theta3	-0,41	1.005,00	7,86	-1.958	-11,35	1.963	0,78%	0,66
theta4	-0,17	1.006,00	7,78	-1.999	5,17	1.945	0,77%	0,85
theta5	-314,70	1.321,00	10,29	-3.275	-253,00	1.948	0,78%	0,00
theta6	12,23	1.002,00	8,21	-1.943	17,75	1.972	0,82%	204.843,18
theta7	3,24	1.007,00	8,80	-1.987	13,76	1.957	0,87%	25,41
theta8	5,25	1.006,00	7,69	-1.956	8,86	1.961	0,76%	189,81

Table 14.11. Estimation results for the product ownership (deposit)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
theta0	-3,14	987,00	7,79	-1.939	-7,58	1.953	0,79%	
theta1	1,41	997,00	7,93	-1.989	1,76	1.935	0,80%	4,11
theta2	-0,71	1.006,00	7,63	-1.943	-7,21	1.989	0,76%	0,49
theta3	-10,22	997,90	7,78	-1.949	-13,54	1.964	0,78%	0,00
theta4	10,38	1.008,00	7,99	-1.985	23,88	1.947	0,79%	32.208,96
theta5	8,25	1.008,00	8,19	-1.969	12,10	1.991	0,81%	3.827,63
theta6	4,20	997,60	8,04	-1.934	4,46	1.956	0,81%	66,49
theta7	5,18	1.001,00	8,39	-1.955	8,69	1.955	0,84%	177,15
theta8	6,94	1.001,00	7,65	-1.946	8,64	1.954	0,76%	1.027,62

Table 14.12. Estimation results for the product ownership (investment fund)

Parameter	**Mean**	**S.D.**	**MC error**	**2,50%**	**median**	**97,50%**	**MC error/S.D.**	**e^mean**
theta0	-1,29	992,90	8,09	-1.938	-5,92	1.951	0,81%	
theta1	2,85	1.000,00	8,00	-1.962	-0,77	1.937	0,80%	17,27
theta2	-7,64	1.012,00	7,77	-1.978	-11,66	1.994	0,77%	0,00
theta3	-7,67	1.007,00	8,04	-1.978	-12,55	1.962	0,80%	0,00
theta4	6,08	1.015,00	7,93	-1.993	10,28	1.948	0,78%	437,47
theta5	-355,50	1.390,00	10,50	-3.648	-273,40	1.965	0,76%	0,00
theta6	11,80	1.013,00	8,42	-1.962	9,99	1.982	0,83%	133.252,35
theta7	7,48	1.018,00	8,59	-1.967	12,22	1.980	0,84%	1.775,79
theta8	10,84	1.003,00	8,11	-1.929	15,63	1.958	0,81%	51.021,38

Table 14.13. Estimation results for the product ownership (pension plan)

Parameter	**Mean**	**S.D.**	**MC error**	**2,50%**	**median**	**97,50%**	**MC error/S.D.**	**e^mean**
theta8	10,84	1.003,00	8,11	-1.929	15,63	1.958	0,81%	
theta1	6,31	995,90	7,88	-1.976	11,38	1.950	0,79%	548,40
theta2	6,91	1.007,00	7,38	-1.934	3,75	1.988	0,73%	1.002,25
theta3	-5,27	997,40	7,47	-1.958	-10,21	1.954	0,75%	0,01
theta4	3,56	1.011,00	7,93	-1.982	22,51	1.941	0,78%	35,02
theta5	4,57	1.005,00	8,53	-1.952	3,10	1.982	0,85%	96,06
theta6	12,04	1.001,00	8,12	-1.930	21,22	1.965	0,81%	169.396,94
theta7	8,73	1.017,00	8,63	-1.970	10,38	1.989	0,85%	6.210,52
theta8	13,16	1.003,00	8,11	-1.934	18,60	1.974	0,81%	519.176,92

Table 14.14. Estimation results for the product ownership (securities)

Parameter	**Mean**	**S.D.**	**MC error**	**2,50%**	**median**	**97,50%**	**MC error/S.D.**	**e^mean**
theta0	-12,71	994,90	7,53	-1.938	-15,17	1.970	0,76%	
theta1	7,37	994,90	7,77	-1.965	9,09	1.954	0,78%	1.582,88
theta2	1,18	1.005,00	7,65	-1.949	1,65	1.975	0,76%	3,27
theta3	-6,46	1.004,00	7,70	-1.965	-14,30	1.973	0,77%	0,00
theta4	8,31	1.005,00	7,93	-1.965	9,81	1.952	0,79%	4.064,31
theta5	6,37	995,50	8,46	-1.942	14,53	1.947	0,85%	583,47
theta6	15,00	1.003,00	8,33	-1.930	26,93	1.978	0,83%	3.269.017,37
theta7	9,68	1.012,00	8,90	-1.960	17,82	1.972	0,88%	15.914,72
theta8	4,79	1.006,00	7,87	-1.949	10,83	1.970	0,78%	120,18

Table 14.15. Estimation results for the product ownership (consumer loan)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
theta0	2,11	988,70	7,64	-1.924	1,53	1.954	0,77%	
theta1	9,07	997,10	7,59	-1.982	13,01	1.945	0,76%	8.655,93
theta2	-2,19	1.006,00	7,64	-1.941	-11,07	1.970	0,76%	0,11
theta3	-2,37	995,90	7,65	-1.961	-0,05	1.954	0,77%	0,09
theta4	6,00	1.006,00	7,74	-1.997	14,63	1.961	0,77%	402,22
theta5	-3,45	997,20	8,58	-1.949	5,44	1.951	0,86%	0,03
theta6	9,43	1.004,00	8,06	-1.949	14,91	1.964	0,80%	12.481,46
theta7	-1,74	1.009,00	8,40	-1.988	-1,60	1.960	0,83%	0,18
theta8	4,90	1.000,00	8,19	-1.959	17,91	1.950	0,82%	134,42

Table 14.16. Estimation results for the product ownership (micro consumer loan)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
theta0	-9,16	990,90	7,62	-1.955	-8,73	1.935	0,77%	
theta1	8,00	991,20	7,52	-1.964	12,84	1.946	0,76%	2.992,91
theta2	-8,14	1.003,00	7,50	-1.944	-15,22	1.951	0,75%	0,00
theta3	-1,54	1.001,00	7,62	-1.961	-6,71	1.981	0,76%	0,21
theta4	10,40	1.013,00	8,43	-1.986	17,64	1.982	0,83%	32.859,63
theta5	7,97	1.005,00	8,42	-1.954	10,94	1.944	0,84%	2.889,97
theta6	15,32	998,00	8,25	-1.922	26,49	1.968	0,83%	4.501.854,59
theta7	5,24	1.015,00	8,84	-1.990	14,65	1.959	0,87%	188,86
theta8	5,68	999,90	7,94	-1.957	12,23	1.929	0,79%	292,95

Table 14.17. Estimation results for the product ownership (mortgage)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
theta0	-14,87	991,50	7,82	-1.946	-11,63	1.943	0,79%	
theta1	8,05	1.001,00	7,83	-1.956	-3,73	1.971	0,78%	3.146,36
theta2	5,33	998,40	7,43	-1.926	7,72	1.973	0,74%	205,61
theta3	-0,41	994,70	7,60	-1.943	-13,03	1.968	0,76%	0,66
theta4	10,39	1.007,00	7,69	-1.976	17,35	1.965	0,76%	32.532,67
theta5	4,37	998,50	8,14	-1.969	4,48	1.992	0,82%	79,04
theta6	6,42	995,40	8,27	-1.915	6,24	1.954	0,83%	615,85
theta7	6,94	1.018,00	8,57	-2.002	14,50	1.970	0,84%	1.035,87
theta8	10,24	1.002,00	7,97	-1.942	16,17	1.978	0,80%	28.001,13

Table 14.18. Estimation results for the product ownership (credit)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
theta0	-2,80	989,70	7,76	-1.922	2,20	1.941	0,78%	
theta1	4,07	995,40	7,56	-1.983	6,64	1.934	0,76%	58,73
theta2	-1,33	1.005,00	7,53	-1.949	-0,45	1.974	0,75%	0,27
theta3	-4,73	998,60	7,59	-1.965	-7,36	1.949	0,76%	0,01
theta4	5,47	1.005,00	7,87	-1.973	20,27	1.958	0,78%	238,41
theta5	5,38	1.005,00	8,32	-1.945	6,20	1.998	0,83%	217,24
theta6	17,01	992,80	8,15	-1.937	28,87	1.960	0,82%	24.397.714,06
theta7	4,20	1.011,00	8,66	-1.965	8,87	1.962	0,86%	66,42
theta8	3,30	1.007,00	7,79	-1.949	6,41	1.964	0,77%	27,19

(2) Monitor the **trace plots**. They are the plots of the iterations versus the generated values. If all values are within a zone without strong periodicities and (especially) tendencies, then we can assume convergence (Ntzoufras, 2009 p.41), as is our case (see Figures 10 in Appendix C1).

(3) Monitor the **history plots**. History option draws a full trace plot of all stored values, while the trace plot provides an online plot of the values generated (Ntzoufras, 2009 p.119). Regarding history plots, if no patterns or irregularities are observed, convergence can be assumed (Ntzoufras, 2009 p.129), such is our case. For more details see Figures 11 in Appendix C1.

(4) Monitor the **quantiles plots**. This is a plot of the evolution for the median and the 2,5% and 97,5% percentiles for each iteration of the algorithm (Ntzoufras, 2009 p.119). This plot indicates if the requested quantiles have been stabilised, implying that the algorithm has converged in terms of the parameters of the model (Ntzoufras, 2009 p.130). Our quantiles are stabilised, so convergence can be assumed (see Figures 12 in Appendix C1).

(5) Run multiple chains in order to perform **Gelman-Rubin convergence diagnostic** (Gelman and Rubin, 1992). Another tactic to check convergence implies running multiple (two or more) chains, with different starting points (inits). When we observe that the lines of different chains mix or cross in trace plots, then convergence is ensured. More specifically, Gelman and Rubin (1992) proposed a convergence test based on two or more parallel chains, each started from different initial values, which are overdispersed with respect to the true posterior distribution. Then, an ANOVA-type diagnostic test is implemented by calculating and comparing the between-sample and the within-sample variability (i.e., intersample and intrasample variability). Therefore,

in essence, their method is based on a comparison of the within and between chain variances for each variable. This comparison is used to estimate the factor by which the scale parameter of the marginal posterior distribution of each parameter might be reduced (also called the shrink factor) if the chain were run to infinity. The statistic R can be estimated by the following formula, where k is the number of generated samples/chains, T' is the number of iterations kept in each sample/chain, BSS/T' is the variance of the posterior mean values over all generated samples/chains (between-sample variance) and WSS is the mean of the variances within each sample (within-sample variability):

$$\hat{R} = \frac{\hat{V}}{WSS} = \frac{T'-1}{T'} = \frac{BSS/T'}{WSS}\frac{k+1}{k}$$

The pooled posterior variance estimate is given by:

$$\hat{V} = \frac{T'-1}{T'}WSS + \frac{BSS}{T'}\frac{k+1}{k}$$

When convergence is achieved and the size of the generated data is large, then $\hat{R} \to 1$. Brooks and Gelman (1998) adopted a corrected version of this statistic given by the following formula, where d represents the estimated degrees of freedom for the pooled posterior variance estimate $\hat{V}$ (for more details see Brooks and Gelman, 1998 sec.3):

$$\hat{R}_c = \frac{d+3}{d+1}\hat{R}$$

Therefore, the Gelman-Rubin convergence diagnostic reports the 50% and 97,5% percentiles of the sampling distribution for the previously mentioned shrink factor, where these percentiles are estimated from the second half of each chain. If both percentiles are approximately equal to 1, effective convergence may be diagnosed. In other words, the samples from the second half of each chain may be assumed to have arisen from the same stationary distribution. Empirical results showed that for all parameters of the model, both percentiles are close to 1 indicating convergence of the estimates. Graphically all lines are stabilised (see Figures 13 in Appendix C1), which implies that convergence has been achieved (Ntzoufras, 2009 p.144).

6.4.2.2. Product usage model

The proposed hierarchical Bayes Poisson model for product usage is also estimated by means of a Markov Chain Monte Carlo (MCMC) technique based on Gibbs sampling to explore the posterior distribution of the parameters of interest using WinBUGS version 1.4.3. This task is accomplished by entering the evidence provided by the observed results of the dependent variable (product usage) and updating the prior distributions by means of the Bayes' Theorem using a MCMC-based procedure (for more details about the model code see Appendix A2).

500 Iterations are chosen as a *burn-in period* after which another 20.500 iterations that are run to obtain the *posterior distribution* of the parameters of interest (i.e., $\beta_0, \dots, \beta_8$). In order to verify parameter convergence, we use different diagnostics, as we have shown in the previous section about product ownership model: (1) Monte Carlo error (MC error) and autocorrelations; (2) trace plots; (3) history plots; (4) quantiles plots; and (5) Gelman-Rubin diagnostic.

(1) **Monte Carlo error (MC error)** and the **autocorrelations** (see Tables 15 enclosed below and Figures 14 in Appendix C2). The small values of MC error are lower than the 1%-5% of the corresponding posterior standard deviation. Therefore, the quantity of interest is calculated with precision. Additionally, if autocorrelations are low, convergence has been obtained in a relative low number of iterations (Ntzoufras, 2009 p.120). For some of the parameters estimated, autocorrelations are high, which implies that to achieve the desired convergence more iterations should be run. For these models we have proved with more than 300.000 iterations but results related to problematic parameters do not improve.

Table 15.1. Estimation results for the product usage (stock capital)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
beta0	-4,81	0,78	0,04	-6,47	-4,74	-3,45	5,38%	
beta1	0,13	0,03	1,76E-03	0,07	0,13	0,20	5,38%	-
beta2	**-1,63**	**0,03**	**3,96E-04**	**-1,69**	**-1,64**	**-1,56**	**1,20%**	**0,20**
beta3	-4,27E-03	0,03	3,28E-04	-0,06	-3,30E-03	0,04	1,30%	1,00
beta4	**0,35**	**0,10**	**1,04E-03**	**0,16**	**0,35**	**0,56**	**1,06%**	**1,42**
beta5	**2,14E-05**	**6,57E-06**	**4,15E-08**	**1,12E-05**	**2,03E-05**	**3,72E-05**	**0,63%**	**1,00**
beta6	2,39E-06	1,67E-06	1,04E-08	-7,41E-07	2,29E-06	5,94E-06	0,62%	1,00
beta7	0,10	0,10	7,63E-04	-0,11	0,10	0,30	0,74%	1,10
beta8	**19,83**	**0,16**	**1,39E-03**	**19,39**	**19,88**	**20,00**	**0,84%**	**409.316.805,60**

Table 15.2. Estimation results for the product usage (credit card)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
beta0	-8,28	2,46	0,14	-15,50	-7,85	-4,85	5,75%	
beta1	0,18	0,10	5,91E-03	0,04	0,17	0,48	5,75%	-
beta2	**-1,36**	**0,09**	**1,19E-03**	**-1,52**	**-1,36**	**-1,16**	**1,30%**	**0,26**
beta3	-0,08	0,06	8,15E-04	-0,21	-0,08	0,02	1,39%	0,92
beta4	**0,61**	**0,16**	**1,77E-03**	**0,32**	**0,60**	**0,97**	**1,08%**	**1,85**
beta5	**3,90E-06**	**1,47E-06**	**8,82E-09**	**9,88E-07**	**3,98E-06**	**6,62E-06**	**0,60%**	**1,00**
beta6	-5,12E-07	3,80E-06	2,58E-08	-8,94E-06	-1,36E-07	5,81E-06	0,68%	1,00
beta7	**1,16**	**0,23**	**2,81E-03**	**0,71**	**1,15**	**1,62**	**1,22%**	**3,17**
beta8	**19,50**	**0,48**	**4,04E-03**	**18,21**	**19,64**	**19,99**	**0,84%**	**294.267.566,04**

Table 15.3. Estimation results for the product usage (debit card)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
beta0	**-0,63**	**0,20**	**7,04E-03**	**-1,02**	**-0,63**	**-9,26**	**3,61%**	
beta1	-8,00E-03	8,50E-03	3,06E-04	-0,02	-8,03E-03	0,71	3,60%	0,99
beta2	**-1,69**	**0,04**	**8,50E-04**	**-1,77**	**-1,69**	**-1,60**	**2,23%**	**0,18**
beta3	-0,03	0,03	7,23E-04	-0,09	-0,03	0,03	2,23%	0,97
beta4	-0,19	0,13	2,13E-03	-0,45	-0,19	0,03	1,62%	0,83
beta5	-1,01E-06	8,85E-07	5,03E-09	-2,78E-06	-9,97E-07	7,71E-07	0,57%	1,00
beta6	**-9,15E-06**	**2,23E-06**	**1,46E-08**	**-1,37E-05**	**-9,08E-06**	**-4,73E-06**	**0,66%**	**1,00**
beta7	**0,68**	**0,07**	**5,88E-04**	**0,54**	**0,68**	**0,82**	**0,85%**	**1,97**
beta8	**19,86**	**0,14**	**1,31E-03**	**19,50**	**19,90**	**20,00**	**0,96%**	**421.782.358,16**

Table 15.4. Estimation results for the product usage (saving insurance)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
beta0	-13,42	3,75	0,21	-19,64	-13,38	-6,81	5,71%	
beta1	0,37	0,16	8,96E-03	0,09	0,37	0,63	5,71%	-
beta2	**-1,33**	**0,14**	**1,47E-03**	**-1,56**	**-1,34**	**-1,01**	**1,03%**	**0,27**
beta3	0,02	0,07	7,60E-04	-0,12	0,02	0,15	1,12%	1,02
beta4	0,46	0,23	2,20E-03	-0,04	0,48	0,88	0,94%	1,59
beta5	-2,82E-07	2,69E-06	1,58E-08	-6,65E-06	1,17E-07	3,82E-06	0,59%	1,00
beta6	1,91E-06	4,19E-06	2,52E-08	-7,95E-06	2,48E-06	8,47E-06	0,60%	1,00
beta7	0,23	0,35	2,94E-03	-0,46	0,23	0,92	0,83%	1,26
beta8	**18,96**	**0,95**	**7,90E-03**	**16,47**	**19,23**	**19,97**	**0,83%**	**171.483.909,84**

Table 15.5. Estimation results for the product usage (home insurance)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
beta0	-14,67	3,85	0,23	-19,84	-15,23	-7,07	5,84%	
beta1	0,45	0,16	9,42E-03	0,13	0,47	0,67	5,84%	-
beta2	**-1,44**	**0,08**	**8,36E-04**	**-1,58**	**-1,44**	**-1,27**	**1,08%**	**0,24**
beta3	-0,03	0,05	6,82E-04	-0,14	-0,03	0,07	1,26%	0,97
beta4	**0,48**	**0,17**	**1,56E-03**	**0,14**	**0,48**	**0,80**	**0,93%**	**1,61**
beta5	**1,88E-05**	**1,92E-06**	**1,43E-08**	**1,51E-05**	**1,87E-05**	**2,27E-05**	**0,75%**	**1,00**
beta6	**-1,79E-05**	**1,07E-05**	**6,67E-08**	**-4,22E-05**	**-1,67E-05**	**-6,51E-07**	**0,62%**	**1,00**
beta7	**0,66**	**0,24**	**2,33E-03**	**0,19**	**0,66**	**1,14**	**0,96%**	**1,93**
beta8	**19,58**	**0,40**	**3,35E-03**	**18,48**	**19,70**	**19,99**	**0,83%**	**318.776.248,73**

Table 15.6. Estimation results for the product usage (not linked life insurance)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
beta0	-14,17	3,82	0,22	-19,78	-14,60	-6,47	5,71%	
beta1	0,41	0,16	9,12E-03	0,08	0,43	0,64	5,71%	-
beta2	**-1,20**	**0,15**	**1,83E-03**	**-1,45**	**-1,21**	**-0,87**	**1,24%**	**0,30**
beta3	-0,06	0,09	1,17E-03	-0,26	-0,06	0,10	1,30%	0,94
beta4	**0,59**	**0,23**	**2,23E-03**	**0,12**	**0,60**	**1,04**	**0,96%**	**1,81**
beta5	-2,72E-06	4,11E-06	2,71E-08	-1,25E-05	-2,04E-06	3,34E-06	0,66%	1,00
beta6	**-4,65E-05**	**2,49E-05**	**1,50E-07**	**-1,03E-04**	**-4,35E-05**	**-7,10E-06**	**0,60%**	**1,00**
beta7	0,61	0,36	3,52E-03	-0,09	0,60	1,33	0,98%	1,84
beta8	**18,97**	**0,93**	**7,87E-03**	**16,56**	**19,24**	**19,97**	**0,85%**	**173.207.351,79**

Table 15.7. Estimation results for the product usage (linked life insurance)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
beta0	-8,88	2,12	0,12	-12,89	-8,98	-5,15	5,72%	
beta1	0,26	0,09	5,07E-03	0,10	0,26	0,42	5,72%	-
beta2	**-1,42**	**0,07**	**1,05E-03**	**-1,54**	**-1,42**	**-1,27**	**1,53%**	**0,24**
beta3	**-0,10**	**0,05**	**8,38E-04**	**-0,21**	**-0,10**	**-4,62E-03**	**1,59%**	**0,90**
beta4	**0,57**	**0,14**	**1,55E-03**	**0,29**	**0,56**	**0,85**	**1,11%**	**1,76**
beta5	**1,41E-05**	**1,95E-06**	**1,44E-08**	**1,03E-05**	**1,40E-05**	**1,80E-05**	**0,74%**	**1,00**
beta6	**-5,23E-05**	**1,21E-05**	**7,88E-08**	**-7,65E-05**	**-5,22E-05**	**-2,88E-05**	**0,65%**	**1,00**
beta7	0,21	0,18	1,55E-03	-0,14	0,21	0,57	0,85%	1,24
beta8	**19,65**	**0,33**	**2,91E-03**	**18,77**	**19,75**	**19,99**	**0,88%**	**341.890.134,75**

Table 15.8. Estimation results for the product usage (other insurances)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
beta0	-7,47	2,40	0,12	-12,95	-7,04	-3,88	5,13%	
beta1	0,10	0,10	5,22E-03	-0,05	0,08	0,33	5,13%	-
beta2	**-1,21**	**0,16**	**1,85E-03**	**-1,48**	**-1,23**	**-0,85**	**1,14%**	**0,30**
beta3	4,53E-03	0,08	9,31E-04	-0,16	8,82E-03	0,15	1,18%	1,00
beta4	**0,74**	**0,20**	**1,94E-03**	**0,31**	**0,75**	**1,11**	**0,96%**	**2,09**
beta5	2,96E-06	2,08E-06	1,21E-08	-1,45E-06	3,05E-06	6,94E-06	0,58%	1,00
beta6	**-4,17E-05**	**2,49E-05**	**1,70E-07**	**-1,01E-04**	**-3,79E-05**	**-4,35E-06**	**0,68%**	**1,00**
beta7	0,55	0,41	3,80E-03	-0,26	0,55	1,36	0,92%	1,73
beta8	**18,69**	**1,16**	**0,01**	**15,70**	**19,02**	**19,96**	**0,94%**	**130.907.300,38**

Table 15.9. Estimation results for the product usage (account)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
beta0	**-12,37**	**4,69**	**0,13**	**-19,28**	**-12,99**	**-1,79**	**2,70%**	
beta1	16,71	2,63	0,14	10,51	17,45	19,87	5,42%	-
beta2	-16,41	2,65	0,14	-19,77	-17,14	-10,03	5,33%	-
beta3	-0,59	1,13	0,04	-2,96	-0,60	1,59	3,44%	0,55
beta4	-1,52	11,03	0,07	-19,11	-2,00	18,28	0,66%	0,22
beta5	6,06	6,66	0,29	-2,00E-03	3,49	19,19	4,41%	430,09
beta6	**7,25**	**5,82**	**0,23**	**1,05**	**3,96**	**19,20**	**3,87%**	**1.402,48**
beta7	7,67	7,37	0,08	-5,71	7,93	19,41	1,10%	2.149,52
beta8	3,58	4,11	0,09	-4,41	3,73	12,01	2,07%	35,84

Table 15.10. Estimation results for the product usage (home loan)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
beta0	-15,49	3,14	0,14	-19,83	-15,90	-8,66	4,44%	
beta1	0,16	0,14	6,62E-03	-0,07	0,14	0,46	4,58%	-
beta2	**-0,80**	**0,35**	**5,80E-03**	**-1,47**	**-0,80**	**-0,14**	**1,66%**	**0,45**
beta3	-0,32	0,30	5,24E-03	-0,97	-0,29	0,18	1,74%	0,73
beta4	-0,08	0,81	8,64E-03	-1,82	-8,07E-03	1,27	1,07%	0,92
beta5	**-10,18**	**5,73**	**0,03**	**-19,51**	**-10,26**	**-0,58**	**0,53%**	**0,00**
beta6	-1,58E-07	1,32E-05	1,02E-07	-3,14E-05	1,95E-06	2,10E-05	0,77%	1,00
beta7	**7,34**	**3,26**	**0,15**	**2,26**	**6,95**	**14,31**	**4,51%**	**-**
beta8	**17,86**	**1,80**	**0,02**	**13,27**	**18,32**	**19,94**	**0,87%**	**57.082.034,91**

Table 15.11. Estimation results for the product usage (deposit)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
beta0	-15,71	3,51	0,21	-19,85	-16,87	-7,27	5,85%	
beta1	0,49	0,15	8,58E-03	0,14	0,54	0,67	5,85%	-
beta2	**-1,54**	**0,06**	**8,29E-04**	**-1,66**	**-1,54**	**-1,41**	**1,29%**	**0,21**
beta3	-0,07	0,05	7,22E-04	-0,17	-0,07	0,03	1,39%	0,93
beta4	**0,77**	**0,19**	**2,41E-03**	**0,44**	**0,75**	**1,18**	**1,27%**	**2,16**
beta5	**-5,36E-05**	**9,42E-06**	**6,54E-08**	**-7,46E-05**	**-5,29E-05**	**-3,71E-05**	**0,69%**	**1,00**
beta6	**2,14E-04**	**1,68E-05**	**1,92E-07**	**1,81E-04**	**2,14E-04**	**2,47E-04**	**1,14%**	**1,00**
beta7	-0,08	0,23	1,94E-03	-0,53	-0,07	0,35	0,86%	0,93
beta8	**19,74**	**0,26**	**2,27E-03**	**19,04**	**19,81**	**19,99**	**0,88%**	**374.087.393,30**

Table 15.12. Estimation results for the product usage (investment fund)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
beta0	**-13,93**	**3,87**	**0,19**	**-19,71**	**-14,30**	**-6,38**	**4,87%**	
beta1	0,25	0,27	0,01	-0,05	0,26	0,53	4,79%	1,28
beta2	**-0,85**	**0,34**	**4,55E-03**	**-1,46**	**-0,86**	**-0,20**	**1,33%**	**0,43**
beta3	-0,25	0,62	0,03	-0,76	-0,20	0,16	4,22%	0,78
beta4	0,62	0,85	0,03	-0,78	0,64	1,75	2,98%	1,86
beta5	**-10,21**	**5,77**	**0,03**	**-19,51**	**-10,31**	**-0,53**	**0,54%**	**0,00**
beta6	1,91E-05	2,84E-04	1,20E-05	-8,96E-06	8,11E-06	1,93E-05	4,21%	1,00
beta7	**3,01**	**1,54**	**0,05**	**0,81**	**2,82**	**6,08**	**3,06%**	**20,23**
beta8	**17,60**	**1,96**	**0,02**	**12,78**	**18,09**	**19,93**	**0,81%**	**44.013.193,53**

Table 15.13. Estimation results for the product usage (pension plan)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
beta0	-7,02	1,99	0,10	-11,77	-6,76	-3,99	5,25%	
beta1	0,10	0,08	4,37E-03	-0,03	0,09	0,30	5,25%	-
beta2	**-1,03**	**0,17**	**1,88E-03**	**-1,34**	**-1,03**	**-0,68**	**1,10%**	**0,36**
beta3	-0,06	0,08	7,94E-04	-0,22	-0,05	0,08	1,04%	0,94
beta4	**0,53**	**0,22**	**1,93E-03**	**0,07**	**0,54**	**0,94**	**0,88%**	**1,70**
beta5	1,49E-06	1,66E-06	9,00E-09	-2,18E-06	1,65E-06	4,35E-06	0,54%	1,00
beta6	**6,59E-06**	**2,46E-06**	**1,57E-08**	**1,18E-06**	**6,81E-06**	**1,08E-05**	**0,64%**	**1,00**
beta7	0,61	0,33	3,32E-03	-0,04	0,60	1,25	1,01%	1,83
beta8	**18,18**	**1,42**	**0,01**	**14,79**	**18,49**	**19,94**	**0,89%**	**78.609.255,11**

Table 15.14. Estimation results for product usage (securities)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
beta0	-13,66	4,13	0,22	-19,72	-14,07	-5,55	5,37%	
beta1	0,31	0,17	9,34E-03	-0,03	0,33	0,58	5,38%	-
beta2	**-1,00**	**0,24**	**2,56E-03**	**-1,40**	**-1,02**	**-0,49**	**1,08%**	**0,37**
beta3	-0,11	0,14	1,62E-03	-0,42	-0,09	0,13	1,15%	0,90
beta4	0,58	0,36	3,22E-03	-0,23	0,61	1,22	0,89%	1,78
beta5	-3,62E-06	5,77E-06	3,68E-08	-1,80E-05	-2,43E-06	4,08E-06	0,64%	1,00
beta6	7,86E-06	3,95E-06	2,88E-08	-1,22E-06	8,31E-06	1,43E-05	0,73%	1,00
beta7	1,09	0,65	7,21E-03	-0,12	1,08	2,42	1,12%	2,98
beta8	**18,16**	**1,57**	**0,01**	**14,15**	**18,57**	**19,95**	**0,79%**	**77.052.687,57**

Table 15.15. Estimation results for the product usage (consumer loan)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
beta0	-14,73	3,50	0,20	-19,77	-15,08	-7,35	5,75%	
beta1	0,44	0,48	0,03	0,15	0,48	0,68	5,74%	-
beta2	**-0,97**	**0,21**	**7,00E-03**	**-1,28**	**-0,99**	**-0,63**	**3,34%**	**0,38**
beta3	**-0,23**	**0,23**	**0,01**	**-0,44**	**-0,21**	**-0,05**	**4,88%**	**0,79**
beta4	**0,75**	**0,67**	**0,01**	**0,35**	**0,74**	**1,20**	**1,65%**	**2,12**
beta5	1,54E-04	2,26E-03	1,27E-04	-1,10E-06	2,35E-06	5,50E-06	5,62%	1,00
beta6	**9,99E-03**	**0,16**	**9,03E-03**	**-3,54E-05**	**-1,58E-05**	**-1,17E-06**	**5,76%**	**1,01**
beta7	0,33	1,33	0,07	-0,28	0,24	0,78	5,39%	1,39
beta8	**18,61**	**1,13**	**0,01**	**15,85**	**18,88**	**19,96**	**0,92%**	**120.842.668,84**

Table 15.16. Estimation results for the product usage (micro consumer loan)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
beta0	-14,31	3,87	0,21	-19,78	-14,99	-6,17	5,50%	
beta1	0,40	0,16	8,93E-03	0,05	0,43	0,63	5,49%	-
beta2	**-0,90**	**0,25**	**3,92E-03**	**-1,32**	**-0,92**	**-0,36**	**1,57%**	**0,41**
beta3	-0,16	0,16	2,96E-03	-0,55	-0,14	0,09	1,84%	0,85
beta4	**1,51**	**0,38**	**6,40E-03**	**0,89**	**1,46**	**2,42**	**1,68%**	**4,54**
beta5	**-6,95E-05**	**5,99E-05**	**4,15E-07**	**-2,23E-04**	**-5,41E-05**	**-7,57E-07**	**0,69%**	**1,00**
beta6	**-4,07E-04**	**1,99E-04**	**1,30E-06**	**-8,72E-04**	**-3,77E-04**	**-1,05E-04**	**0,66%**	**1,00**
beta7	-0,07	0,55	4,15E-03	-1,17	-0,06	0,99	0,75%	0,93
beta8	**18,16**	**1,56**	**0,01**	**14,25**	**18,57**	**19,94**	**0,85%**	**77.052.687,57**

Table 15.17. Estimation results for the product usage (mortgage)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
beta0	-11,71	5,42	0,24	-19,64	-12,08	0,37	4,40%	
beta1	-3,06	6,74	0,40	-19,88	0,21	0,60	5,97%	0,05
beta2	**-4,79**	**7,35**	**0,44**	**-19,86**	**-0,92**	**-0,19**	**5,93%**	**0,01**
beta3	-0,56	5,12	0,30	-17,41	-0,77	10,63	5,84%	0,57
beta4	-1,03	6,05	0,25	-18,21	0,74	9,11	4,19%	0,36
beta5	**0,03**	**0,09**	**5,61E-03**	**6,24E-05**	**1,23E-04**	**0,35**	**5,97%**	**1,03**
beta6	**-0,06**	**0,18**	**0,01**	**-0,66**	**-1,93E-04**	**-5,20E-05**	**5,97%**	**0,95**
beta7	-1,26	6,07	0,29	-18,39	0,87	3,59	4,77%	0,28
beta8	**17,89**	**4,18**	**0,18**	**4,72**	**19,06**	**19,97**	**4,22%**	**58.820.441,68**

Table 15.18. Estimation results for the product usage (credit)

Parameter	Mean	S.D.	MC error	2,50%	median	97,50%	MC error/S.D.	e^mean
beta0	-11,44	6,56	0,26	-19,61	-12,52	7,65	3,90%	
beta1	-4,23	8,04	0,48	-19,88	0,24	0,60	5,95%	0,01
beta2	**-4,49**	**6,98**	**0,40**	**-19,65**	**-0,84**	**-0,09**	**5,71%**	**0,01**
beta3	-1,29	6,15	0,34	-17,77	-0,38	16,65	5,59%	0,28
beta4	0,49	6,33	0,19	-16,96	0,79	17,96	3,00%	1,64
beta5	1,66	3,18	0,19	-2,26E-06	6,62E-06	9,56	5,90%	5,24
beta6	**-2,01**	**4,16**	**0,25**	**-13,67**	**-5,97E-04**	**-8,20E-05**	**5,96%**	**0,13**
beta7	0,52	5,95	0,16	-14,83	-0,01	17,20	2,61%	1,68
beta8	15,80	6,27	0,25	-5,81	17,90	19,93	3,92%	7.275.331,96

(2) **Trace plots**. If we see Figures 15 in Appendix C2, we can appreciate that most of the parameter values are within a zone without strong periodicities and tendencies and then we can assume convergence (Ntzoufras, 2009 p.41). For those parameters whose blue chain and red chain are not crossed (as you can see in Figures 15), convergence could not be assumed.

(3) **History plots**. As you can see in Figures 16 (enclosed in Appendix C2), for most of the parameter values no patterns or irregularities are observed and then convergence can be assumed for these parameters (Ntzoufras, 2009 p.129). On the contrary, for some of them we can appreciate patters and irregularities (as you can see in Figures 16), so convergence could not be assumed for these parameters.

(4) **Quantiles plots**. As you can see in Figures 17 (in Appendix C2), most of the requested quantiles are stabilised and then the algorithm has converged in terms of these parameters of the model (Ntzoufras, 2009 p.130). For those parameters whose quantiles are not stabilised, convergence could not be assumed.

(5) Run multiple chains in order to perform **Gelman-Rubin convergence diagnostic** (Gelman and Rubin, 1992). If the lines of different chains mix or cross in trace plots, convergence has been achieved. More specifically, if we perform the **Gelman-Rubin convergence diagnostic** and both percentiles are approximately equal to 1, effective convergence may be diagnosed. This is our case for most of our parameter values, but as we have indicated in the previously shown convergence diagnostics, some other parameters have not achieved convergence. Graphically, most lines are stabilised (see Figures 18 in Appendix C2), which implies that convergence has been achieved for these parameters (Ntzoufras, 2009 p.144).

6.4.2.3. Contribution margin model

The proposed hierarchical Bayes normal model for contribution margin is also estimated by means of a Markov Chain Monte Carlo (MCMC) technique based on Gibbs sampling to explore the posterior distribution of the parameters of interest using WinBUGS version 1.4.3. This task is accomplished by entering the evidence provided by the observed results of the dependent variable (contribution margin) and updating the prior distributions by means of the Bayes' Theorem using a MCMC-based procedure (for more details about the model code see Appendix A3).

500 Iterations are chosen as a *burn-in period* after which another 50.500 iterations that are run to obtain the *posterior distribution* of the parameters of interest (i.e., $\rho_0, \rho, \tau, \sigma$). In order to verify parameter convergence, we use different diagnostics, as we have shown in previous sections about product ownership and product usage models: (1) Monte Carlo error (MC error) and autocorrelations; (2) trace plots; (3) history plots; (4) quantiles plots; and (5) Gelman-Rubin diagnostic.

(1) **Monte Carlo error (MC error)** and the **autocorrelations** (see Table 16 enclosed below and Figure 19 in Appendix C3). The small values of MC error are lower than the 1% of the corresponding posterior standard deviation. Therefore, the quantity of interest is calculated with precision. Additionally, autocorrelations are low, so convergence has been obtained in a relative low number of iteration (Ntzoufras, 2009 p.120).

Table 16. Estimation results for the contribution margin

Parameter	**Mean**	**S.D.**	**MC error**	**2,50%**	**median**	**97,50%**	**MC error/S.D.**
rho0	-7,945	9,335	0,02746	-26,22	-7,917	10,39	0,29%
rho[1]	0,2939	0,4008	0,001226	-0,4917	0,2923	1,083	0,31%
rho[2]	-0,9318	1,035	0,003321	-2,968	-0,932	1,088	0,32%
rho[3]	-1,287	0,9909	0,00307	-3,229	-1,287	0,6617	0,31%
rho[4]	**-15,68**	**7,458**	**0,0247**	**-30,28**	**-15,66**	**-1,046**	**0,33%**
rho[5]	**0,004362**	**6,842E-05**	**2,208E-07**	**0,004227**	**0,004362**	**0,004497**	**0,32%**
rho[6]	**-0,005483**	**9,801E-05**	**3,238E-07**	**-0,005675**	**-0,005483**	**-0,00529**	**0,33%**
rho[7]	-8,059	4,111	0,01329	-16,14	-8,059	0,02075	0,32%
rho[8]	**26,69**	**1,517**	**0,00478**	**23,72**	**26,68**	**29,67**	**0,32%**
rho[9]	**0,7322**	**0,004009**	**1,319E-05**	**0,7243**	**0,7322**	**0,7401**	**0,33%**
sigma	236	1,311	0,004084	233,4	236	238,6	0,31%
tau	0,00001796	1,996E-07	6,211E-10	0,00001757	0,00001796	0,00001835	0,31%

(2) Monitor the **trace plots**. They are the plots of the iterations versus the generated values. All values are within a zone without strong periodicities and (especially) tendencies, then we can assume convergence (Ntzoufras, 2009 p.41), as is our case (see Figure 20 in Appendix C3).

(3) **History plots**. As you can see in Figure 21 (enclosed in Appendix C3), no patterns or irregularities are observed and then convergence can be assumed (Ntzoufras, 2009 p.129).

(4) **Quantiles plots**. As you can see in Figure 22 (in Appendix C3), the requested quantiles are stabilised and then the algorithm has converged in terms of the parameters of the model (Ntzoufras, 2009 p.130).

(5) Run multiple chains in order to perform **Gelman-Rubin convergence diagnostic** (Gelman and Rubin, 1992). The lines of different chains mix or cross in trace plots and convergence has been achieved. More specifically, empirical results show that for all parameters of the model, both percentiles are close to 1 indicating convergence of the estimates. Graphically all lines are stabilised (see Figure 23 in Appendix C3), which implies that convergence has been achieved (Ntzoufras, 2009 p.144).

6.4.3. Coefficients estimates

6.4.3.1. Product ownership model

Tables 14 shows some summary statistics for the posterior distributions of the vectors of the parameters (i.e., $\theta_0, \dots, \theta_8$) for the model after a total of 44.000 iterations of the MCMC chain.

$$O_{ij,t} =$$
$$\theta_0 + \theta_1 length\ of\ the\ relationship_i + \theta_2 purchase\ recency_{i,t} +$$
$$\theta_3 cancellation\ recency_{i,t} + \theta_4 crossbuying_{i,t} + \theta_5 average\ monthly\ assets_{i,t} +$$
$$\theta_6 average\ monthly\ liabilities_{i,t} + \theta_7 adoption\ of\ online\ banking_{i,t} +$$
$$\theta_8 one\ period\ lagged\ product\ usage_{ij,t-1}.$$

Interpretation of GLM's coefficients is equivalent to the corresponding interpretation of the parameters in usual normal regression models. Thus, interest lies in (1) whether the effect of each covariate x_k is important (significant in statistical terms) for the prediction or description of the response variable *Y*, (2) the type of association between Y and each covariate x_k (positive, negative, linear or other) and (3) the magnitude of the effect of each x_k on *Y* (Ntzoufras, 2009 p. 238).

(1) Concerning the importance of the effect, we can simply monitor the posterior distribution of the parameters and report whether the zero value is away from its center. In other words, there is a close relationship between credible intervals (C.I.) and significance tests. Specifically, if a statistic is significantly different from 0 at the 0,05 level, then the 95% C.I. will not contain 0. On the contrary, if the 95% C.I. contains 0, then the effect will not be significant at the 0,05 level. All values in the C.I. are plausible values for the parameter, whereas values outside the interval are rejected as plausible values for the parameter. Whenever an effect is significant, all values in the C.I. will be on the same side of zero (either all positive or all negative). Therefore, a significant finding allows the researcher to specify the direction of the effect through (2) and the interpretation of the parameter through (3).

(2) The type of the association (positive or negative) is simply indicated by the corresponding sign of the posterior summaries for each parameter as is common in regression models (i.e., column called mean in Tables about estimation results; for product ownership models see Tables 14).

(3) Concerning the interpretation of the model parameters, interest lies in quantifying the effect of each covariate (x_k) on the corresponding parameter of interest of the response variable Y, here $O_{ij,t}$ (usually on the mean of $O_{ij,t}$). Although this is straightforward in normal regression models (such is our case with the contribution margin model), since the canonical link is used and, therefore, the effect of each covariate is linear to the mean of $CM_{i,t}$, it is slightly more complicated in GLMs and depends on the form of the link function (Ntzoufras, 2009 p.238). Is for this reason that we enclose the column e^{mean} in Tables 14. The coefficients θ measure the partial impact of each covariate (x_k) on $\log(\frac{p_i}{1-p_i})$, and consequently, e^{θ} measures the partial impact of each covariate on the odds ratio.

For the product ownership models all covariates appear to have a non-significant impact on all the product ownership variables considered (at the 5% significance level), because in all C.I.'s zero is included. Therefore, a prediction derived from these Bayesian regressions is quite inaccurate.

According to the research question defined in Chapter 1: which drivers of CLV have more potential to predict components of CLV? All the covariates that we take into account in order to predict product ownership do not seem to have a significant impact over this product ownership (see Tables 14). We indicate some comments about the estimations performed below, according with each product considered:

- For stock capital: any of the drivers appear to have a significant impact over the ownership of stock capital because all C.I.'s are quite wide, including 0 (as you can see in Table 14.1) and predictions about the ownership of stock capital using this model are inaccurate (for more details see section 6.4.4. regarding model validation).

- For credit card: any of the variables appear to have a significant impact over the ownership of credit card. This is because all C.I.'s of the parameters of the models are quite wide, including zero. Therefore, predictions about ownership of credit card are inaccurate (for more details see section 6.4.4.).

- For debit card: once more any of the variables appear to have a significant impact over the ownership of debit card. This is because all C.I.'s of the parameters of the models are quite wide, including zero. Therefore, predictions about ownership of debit card are also inaccurate (for more details see section 6.4.4.).

- For saving insurance: again any of the variables appear to have a significant impact over the ownership of saving insurance. This is because all C.I.'s of the parameters of the models include zero. Therefore, predictions about ownership of debit card are inaccurate (for more details see section 6.4.4.).

- For home insurance: in this case we have also obtained C.I.'s that are quite wide (including 0). Predictions in this case are also inaccurate (for more details see section 6.4.4.).

- For not linked life insurance: again we have also obtained C.I.'s that are quite wide (including 0). Predictions in this case are also inaccurate (for more details see section 6.4.4.).

- For linked life insurance: once more predictions about the ownership of linked life insurance are inaccurate for the same reason that we have explained in case of previous products (for more details see section 6.4.4.).

- For other insurances: again predictions are inaccurate (for more details see section 6.4.4.).

- For account: again any of the variables appear to have a significant impact over the ownership of account. Predictions about ownership of account are inaccurate (for more details see section 6.4.4.).

- For home loan: predictions are inaccurate also in this case (for more details see section 6.4.4.).

- For deposit: we have also obtained C.I.'s that are quite wide (including 0). Predictions are also inaccurate (for more details see section 6.4.4.).

- For investment fund: again predictions are inaccurate (for more details see section 6.4.4.).

- For pension plan: again predictions are inaccurate also in this case (for more details see section 6.4.4.).

- For securities: again we have also obtained C.I.'s that are quite wide (including 0). Predictions are also inaccurate (for more details see section 6.4.4.).

- For consumer loan: again any of the variables appear to have a significant impact over the ownership of consumer loan. Therefore, predictions about ownership of this product are inaccurate (for more details see section 6.4.4.).

- For micro consumer loan: again predictions are inaccurate also in this case (for more details see section 6.4.4.).

- For mortgage: predictions are also inaccurate (for more details see section 6.4.4.).

- For credit: once more predictions are inaccurate (for more details see section 6.4.4.).

6.4.3.2. Product usage model

Tables 15 shows some summary statistics for the posterior distributions of the vectors of the parameters (i.e., $\beta_0, \ldots, \beta_8$) for the model after a total of 21.000 iterations of the MCMC chain.

$$U_{ij,t} = \beta_0 + \beta_1 length\ of\ the\ relationship_i + \beta_2 purchase\ recency_{i,t} + \beta_3 cancellation\ recency_{i,t} + \beta_4 crossbuying_{i,t} + \beta_5 average\ monthly\ assets_{i,t} + \beta_6 average\ monthly\ liabilities_{i,t} + \beta_7 adoption\ of\ online\ banking_{i,t} + \beta_8 one\ period\ lagged\ product\ usage_{ij,t-1}.$$

As we have noted in the previous section, interpretation of GLM's coefficients is equivalent to the corresponding interpretation of the parameters in usual normal regression models. Thus, interest lies in (1) whether the effect of each covariate x_k is important (significant in statistical terms) for the prediction or description of the response variable *Y*, (2) the type of association between *Y* and each covariate x_k (positive, negative, linear or other) and (3) the magnitude of the effect of each x_k on *Y* (Ntzoufras, 2009 p. 238).

(1) Concerning the importance of the effect in the product usage models, some covariates appear to have a significant impact on the 18 product usage variables considered (at the 5% significance level), because in some C.I.'s zero is not included. When these C.I.'s do not include zero, using the β coefficients the partial impact of each covariate on the product usage variables can be derived from e^{mean} (last column of Tables 15).

(2) The type of the association (positive or negative) is simply indicated by the corresponding sign of the posterior summaries for each parameter (i.e., column called mean in Tables 15).

(3) Concerning the interpretation of the model parameters, interest lies in quantifying the effect of each covariate (x_k) on the corresponding parameter of interest of the response variable Y, here $U_{ij,t}$ (usually on the mean of $U_{ij,t}$). Although this is straightforward in normal regression models (such is our case with the contribution margin model), since the canonical link is used and, therefore, the effect of each covariate is linear to the mean of $CM_{i,t}$, it is slightly more complicated in GLMs and depends on the form of the link function (Ntzoufras, 2009 p.238). Is for this reason that we enclose the column e^{mean} in Tables 15. The coefficients β measure the partial impact of each covariate (x_k) on $\log(\lambda_i)$, and consequently, e^{β} measures the impact on λ_i

According to the research question defined in Chapter 1: which drivers of CLV have more potential to predict components of CLV? Among the covariates that we take into account in order to predict product usage, for all the products considered the one-period lagged variable of product usage seems to have the largest impact on the product usage. This effect is positive for most of the products considered (except for account and credit), that is, if a customer usages more products in period *(t-1)*, logically will usage more products in period *t*. The remaining coefficients can be interpreted in a similar way, except those coefficients whose convergence could not been achieved, such is the case of length of the relationship (β_1). Therefore, covariates that appear to have a significant impact over usage variables considered (at the 5% significance level) are shown below, according with each product considered (in Tables 15 they are highlighted in bold):

- For stock capital: one-period lagged variable of product usage, cross-buying and average monthly assets (all with a significant positive influence over the usage of stock capital) and purchase recency (with a significant negative effect over the usage of stock capital). More specifically, the expected usage of stock capital significantly increases by the previous usage of this same product (by 1.983%), cross-buying (by 35%) and the amount in average monthly assets (by a small proportion). On the other hand, the expected usage of stock capital significantly decreases by the level of purchase recency or the number of months spent without new acquisitions from the bank (by 163%).

- For credit card: one-period lagged variable of the usage of credit card, adoption of online banking (both with a significant positive influence over the usage of credit card), cross-buying and average monthly assets (both with a significant positive influence) and finally, purchase recency (with a significant negative influence over the usage of credit card). More specifically, the expected usage of credit card significantly increases by the previous usage of this same product (by 1.950%), if the customer uses online banking (by

116%), the level of cross-buying (by 61%) and the amount in average monthly assets (by a small proportion). On the other hand, the expected usage of credit card significantly decreases by the level of purchase recency or the number of months spent without new acquisitions from the bank (by 136%).

- For debit card: one-period lagged variable of product usage and adoption of online banking (both with a significant positive effect over the usage of debit card), average monthly liabilities and purchase recency (all with a significant negative influence over the usage of debit card). More specifically, the expected usage of debit card significantly increases by the previous usage of this same product (by 1.986%) and if the customer uses online banking (by 68%) On the other hand, the expected usage of debit card significantly decreases by the level of purchase recency (by 169%) and the amount in average monthly liabilities (by a small proportion).

- For saving insurance: one-period lagged variable of product usage (with a significant positive influence over the usage of saving insurance) and purchase recency (with a significant negative influence over the usage of saving insurance). More specifically, the expected usage of saving insurance significantly increases by the previous usage of this same product (by 1.896%). On the other hand, the expected usage of saving insurance significantly decreases by the level of purchase recency (by 133%).

- For home insurance: one-period lagged variable of product usage, adoption of online banking, cross-buying and average monthly assets (all with a significant positive influence over the usage of home insurance) and finally, purchase recency and average monthly liabilities (with a significant negative effect over the usage of home insurance). More specifically, the expected usage of home insurance significantly increases by the previous usage of this same product (by 1.958%), if the customer uses online banking (by 66%), the cross-buying (by 48%) and the amount in average monthly assets (by a small proportion). On the other hand, the expected usage of saving insurance significantly decreases by the level of purchase recency (by 144%) and the amount in average monthly liabilities (by a small proportion).

- For not linked life insurance: one-period lagged variable of product usage, cross-buying (both with a significant positive influence over the usage of this product), purchase recency and average monthly liabilities (both with a significant negative influence over the usage of this product). More specifically, the expected usage of not linked life insurance significantly increases by the previous usage of this same product (by 1.897%) and the level of cross-buying (by 59%). On the other hand, the expected usage of not linked life insurance significantly decreases by the level of purchase recency (by 120%) and the amount in average monthly liabilities (by a small proportion).

- For linked life insurance: one-period lagged variable of product usage, cross-buying and average monthly assets (all with a significant positive influence over the usage of this product), purchase recency, cancellation recency and average monthly liabilities (all with a significant negative effect over the usage of linked life insurance). More specifically, the expected usage of linked life insurance significantly increases by the previous usage of this same product (by 1.965%), the level of cross-buying (by 57%) and the amount in average monthly assets (by a small proportion). On the other hand, the expected usage of linked life insurance significantly decreases by the level of purchase recency (by 142%), the level of cancellation recency (by 10%) and the amount in average monthly liabilities (by a small proportion).

- For other insurances: one-period lagged variable of product usage and cross-buying (both with a significant positive influence over the usage of this product), purchase recency and average monthly liabilities (both with a significant negative effect over the usage of other insurances). More specifically, the expected usage of other insurances significantly increases by the previous usage of this same product (by 1.869%) and the level of cross-buying (by 74%). On the other hand, the expected usage of other insurances significantly decreases by the level of purchase recency (by 121%) and the amount in average monthly liabilities (by a small proportion).

- For account: only the predictor average monthly liabilities exerts a significant positive influence over the usage of this product. More specifically, the expected usage of account significantly increases by the amount in average monthly liabilities (by 725%). It is interesting to remark that one-period lagged variable of product usage does not have a significant impact over the usage of account because zero is included in C.I.

- For home loan: one-period lagged variable of product usage and adoption of online banking (both with a significant positive effect over the usage of this product) and average monthly assets and purchase recency (both with a significant negative effect).

More specifically, the expected usage of home loan significantly increases by the previous usage of this same product (by 1.786%) and if the customer uses online banking (by 734%). On the other hand, the expected usage of home loan significantly decreases by the amount in average monthly assets (by 1.018%) and the level of purchase recency (by 80%).

- For deposit: one-period lagged variable of product usage, cross-buying (both with a significant positive impact over the usage of deposits) and average monthly liabilities (also with a significant positive effect); additionally, purchase recency and average monthly assets (both with a significant negative effect). More specifically, the expected usage of deposit significantly increases by the previous usage of this same product (by 1.974%), the level of cross-buying (by 77%) and by the amount in average monthly liabilities (by a small proportion). On the other hand, the expected usage of deposit significantly decreases by the level of purchase recency (by 154%) and the amount in average monthly assets (by a small proportion).

- For investment fund: one-period lagged variable of product usage, adoption of online banking (both with a significant positive influence over the usage of investment fund), average monthly assets and purchase recency (both with a significant negative effect). More specifically, the expected usage of investment fund significantly increases by the previous usage of this same product (by 1.760%) and if the customer uses online banking (by 301%). On the other hand, the expected usage of investment fund significantly decreases by the amount in average monthly assets (by 1.021%) and the level of purchase recency (by 85%).

- For pension plan: one-period lagged variable of product usage, cross-buying, average monthly liabilities (all these predictors exert a significant positive influence over the usage of pension plan) and purchase recency (with a significant negative influence over the usage of pension plan). More specifically, the expected usage of pension plan significantly increases by the previous usage of this same product (by 1.818%), the level of cross-buying (by 53%) and by the amount in average monthly liabilities (by a small proportion). On the other hand, the expected usage of pension plan significantly decreases by the level of purchase recency (by 103%).

- For securities: one-period lagged variable of product usage (with a significant positive effect) and purchase recency (with a significant negative effect). More specifically, the expected usage of securities significantly increases by the previous usage of this same product (by 1.816%). On the other hand, the expected usage of securities significantly decreases by the level of purchase recency (by 100%).

- For consumer loan: one-period lagged variable of product usage, cross-buying and average monthly liabilities (all with a significant positive effect over the usage of this product), cancellation recency and purchase recency (both with a significant negative effect). More specifically, the expected usage of consumer loan significantly increases by the previous usage of this same product (by 1.861%), the level of cross-buying (by 75%) and by the amount in average monthly liabilities (by a small proportion). On the other hand, the expected usage of consumer loan significantly decreases by the level of purchase recency (by 97%) and the level of cancellation recency (by 23%).

- For micro consumer loan: one-period lagged variable of product usage, cross-buying (both exert a significant positive influence over the usage of consumer loan), average monthly assets and average monthly liabilities (both with a significant negative effect) and purchase recency (also with a significant negative effect). More specifically, the expected usage of micro consumer loan significantly increases by the previous usage of this same product (by 1.816%) and cross-buying (by 151%). On the other hand, the expected usage of micro consumer loan significantly decreases by the purchase recency (by 90%) and by the amount in average monthly assets and liabilities (by a small proportion).

- For mortgage: one-period lagged variable of product usage, average monthly assets (both exert a significant positive effect over the usage of this product), average monthly liabilities (significant negative effect) and purchase recency (significant negative effect). More specifically, the expected usage of mortgage significantly increases by the previous usage of this same product (by 1.789%) and by the amount in average monthly assets (by 3%). On the other hand, the expected usage of mortgage significantly decreases by the level of purchase recency (by 479%) and by the amount in average monthly liabilities (by 6%).

- For credit: purchase recency (significant negative effect) and average monthly liabilities (significant negative effect). More specifically, the expected usage of credit significantly decreases by the level of purchase recency (by 449%) and the amount in average monthly liabilities (by 201%).

In the particular cases of credit card, debit card, home insurance, home loan and investment fund, those customers who have adopted online banking, usage more products than the traditional customer population. Therefore, adoption of online banking influences the product usage (for the previously mentioned products), as was stated by Hitt and Frei (2002).

6.4.3.3. Contribution margin model

Table 16 shows some summary statistics for the posterior distributions of the vectors of the parameters (i.e., $\rho_0, \rho, \tau, \sigma$) for the model after a total of 51.000 iterations of the MCMC chain. Using the ρ coefficients, the partial impact of each covariate on the contribution margin can be derived from the column called mean (first column of Table 16).

From the definition of the model we know that:

$$CM_{i,t} =$$
$$\rho_0 + \rho_1 length\ of\ the\ relationship_i + \rho_2 purchase\ recency_{i,t} +$$
$$\rho_3 cancellation\ recency_{i,t} + \rho_4 crossbuying_{i,t} + \rho_5 average\ monthly\ assets_{i,t} +$$
$$\rho_6 average\ monthly\ liabilities_{i,t} + \rho_7 adoption\ of\ online\ banking_{i,t} +$$
$$p_8 total\ quantity\ of\ purchases_{i,t} + \rho_9 one\ period\ lagged\ contribution\ margin_{ij,t-1}.$$

For example, the coefficient $\rho_4 = -15{,}68$ measures the partial impact of $crossbuying$ on CM_{it}. As a result, cross-buying is expected to reduce contribution margin because those customers who purchase different types of banking products (i.e., whose cross-buying variable is higher), will contribute to profits in a small proportion. Maybe because if they owns more products is more likely that they own products with negative contributions to margin. The remaining coefficients can be interpreted in a similar way. For example in case of $total\ quantity\ of\ purchases$, its coefficient is $\rho_8 = 26{,}69$. As a result, total quantity of purchases is expected to increase contribution margin because those customers who own more banking products (they can be different types of banking products or the same product, but the customer owns more than one product), will contribute to profits in a higher proportion.

According to the research question defined in Chapter 1: which drivers of CLV have more potential to predict components of CLV?, the covariates that appear to have a significant impact on contribution margin considered (at the 5% significance level) are (in descending order, that is, arranged from largest to smallest): total quantity of purchases (with a positive influence over contribution margin), cross-buying (with a negative influence over contribution margin), one-period lagged of contribution margin (with a positive influence over contribution margin), average monthly liabilities (with a negative influence over contribution margin), average monthly assets (with a positive influence over contribution margin). In Table 16 they are highlighted in bold. These covariates have a significant effect on contribution margin because in their C.I.'s zero is not included. When these C.I.'s do not include zero, using the ρ coefficients the partial impact of each covariate on the product ownership variables can be derived from the mean (second column of Table 16). On the contrary, adoption of online banking does not have a significant effect over the contribution margin. Therefore, regarding the result that emerges from Hitt and Frei (2002) (i.e., PC banking customers offer a higher contribution margin than the traditional customer population), we can not give support to this idea.

6.4.4. Model validation: comparison between observed values and predictions

Our main objective is to get predictions about product ownership, product usage and contribution margin. We can use the results derived in the implied posterior distributions for the vectors of the parameters (i.e., $\theta_0, \dots, \theta_8, \beta_0, \dots, \beta_8, \rho_0, \rho, \tau, \sigma$) to predict a future occurrence related to product ownership, product usage and contribution margin. We have generated a vector of 1.000 iterations or updates more for the posterior predictive distribution of product ownership, product usage and contribution margin in order to get these predictions. Estimation results about the product ownership models have been inaccurate (except for one product: credit). Therefore, although we are going to show you different comparisons between observed and predicted product ownerships in order to validate the results obtained, we are going to avoid this term (i.e., product ownership) from the CLV formula. The estimation of CLV does not suffer any modification in global terms, because product usage contains the same information as product ownership but in a more detailed way (product ownership is a set of 18 binary variables and product usage is a set of 18 categorical variables).

We validate the results obtained from product ownership and product usage models that have been estimated using *classification matrices*, which represent the levels of predictive accuracy achieved by the models. The measure of predictive accuracy used is the *hit ratio* or the percentage of cases correctly classified (Hair *et al*., 2009 p.266). Moreover, comparisons between the hit ratio and the proportional chance criterion (a measure of random allocation or classification by chance) are also made. The *proportional chance criterion* is used for assessing the hit ratio, in which the average probability of classification is calculated considering all group sizes (Hair *et al*., 2009 p.365). The model performs significantly better when is compared with a classification by chance, because the difference between the two percentages is substantial. We have also checked (with a test for proportions) whether the classification rate for the holdout sample is significantly larger than the percentage due to chance. Despite some results are significant, it should be noted, however, that the z statistic is highly inflated by the size of the sample (n = 1.357 for 12 months leading to 16.284 cases). As a result, the slightest difference is considered to be significant although one cannot, therefore, attach great importance to the significance test.

Table 17. An illustration of the classification matrix

		Predicted 1	Predicted 0	
Actual	**1**	α_{11}	α_{10}	c
	0	α_{01}	α_{00}	d
		a	b	*total simple size*

$$Hit\ ratio = \frac{\alpha_{11} + \alpha_{00}}{total\ sample\ size}$$

$$Proportional\ chance\ criterion = \left(\frac{a}{total\ sample\ size} * \frac{c}{total\ sample\ size}\right) + \left(\frac{b}{total\ sample\ size} * \frac{d}{total\ sample\ size}\right)$$

Finally, in order to examine the overall model fit of contribution margin model (measured using a continuous variable) that has been estimated, we use *Pearson correlation (R)*. It measures the strength and direction of the linear relationship between two variables (in our case observed contribution margin and predicted contribution margin) that is defined as the (sample) covariance of the variables divided by the product of their (sample) standard deviations.

6.4.4.1. Product ownership model

Regarding product ownership models, the overall fit of the models is estimated and the results are shown in Tables 18 (see Appendix D1). For most of the products considered (i.e., 17 products which generates 17 products ownership models), results are not significant ($p > 0{,}01$). Only in case of credit results are significant ($p < 0{,}05$). Therefore, these results confirm that product ownership models generate inaccurate results and justify their exclusion from the overall CLV formula.

Additionally, Figures 24 (see Appendix D1) show the comparison between the observed results related to product ownership for periods *t = 13, ..., 24* (the light grey line) and the estimations provided by our Bernoulli model for the same period of time (the dark grey line) for all the customers and periods in the sample. As one can appreciate, for most of the customers, the proposed Bayesian hierarchical model seems to produce an unacceptable fit over the observed results. This is because of the large differences between the two lines (shown in Figures 24). In general, it seems that our model does not perform well to predict product ownership. Therefore, we are going to avoid this term from the CLV formula. The estimation of CLV does not suffer any modification in global terms, because product usage contains the same information as product ownership but in a more detailed way (product ownership is a set of 18 binary variables and product usage is a set of 18 categorical variables).

6.4.4.2. Product usage model

Regarding product usage models, the overall fit of the models is estimated and the results are shown in Tables 19 (see Appendix D2). For most of the products considered (i.e., 15 products), results are significant. On the contrary, for only three models results are not significant (particularly, for deposit, investment fund and micro consumer loan). Therefore, these results confirm that product usage models generate accurate results and justify their inclusion in the CLV formula.

Additionally, Figures 25 (see Appendix D2) show the comparison between the observed results related to product usage for periods $t = 13, \ldots, 24$ (the light grey line) and the estimations provided by our Poisson model for the same period of time (the dark grey line) for all the customers and periods in the sample. As one can appreciate, for most of the customers, the proposed Bayesian hierarchical model seems to produce an acceptable fit over the observed results. Once more, this is because the differences among the two lines are small and the distribution of the data is quite similar. In general, it seems that our model performs well to predict product usage.

6.4.4.3. Contribution margin model

In order to examine the overall model fit of contribution margin model that has been estimated (in order to compare observed and predicted contribution margins), we use Pearson correlation (R), which is equal to 0,750 ($p < 0{,}01$). This implies that a strong positive relationship exists between predicted contribution margin and observed contribution margin, which indicates that these two variables measure the same concept.

Additionally, Figure 26 (see Appendix D3) shows the comparison between the observed results related to contribution margin for periods $t = 13, \ldots, 24$ (the light grey line) and the estimations provided by our normal model for the same period of time (the dark grey line) for all the customers and periods in the sample. As one can appreciate, for most of the customers, the proposed Bayesian hierarchical model seems to produce a good fit over the observed results. Once more, this is because the differences among the two lines are small and the distribution of the data is quite similar. In general, it seems that our model performs well to predict contribution margin.

6.4.4.4. Customer Lifetime Value

Firstly, we calculate the $\boldsymbol{predicted\ CLV_i}$ using the following formula (where product ownership is avoided). In this formula are included the predictions that we have estimated as components of CLV (i.e., product usage and contribution margin):

$$CLV_i = \sum_{t=1}^{T} \frac{Profit_{i,t}}{(1+d)^t}$$

Where:

CLV_i = lifetime value for customer *i*,

i = index for customers ($1 \leq i \leq I$, I is the total sample size),

t = index for periods of time or months ($1 \leq t \leq T$, T is the end of the calibration or observation time frame; we have used all the 24 months of independent variables and the first 12 months of dependent ones in order to predict the last 12 months of the dependent ones),

$Profit_{i,t}$ = current and future (predicted) contribution margins from the customers of the company, and

d = monthly discount factor, which is the fourth component of CLV.

Additionally, we calculate $Profit_{i,t}$, the main input to get CLV_i. This equation contains the two terms that have been accurately estimated:

$$Profit_{i,t} = \sum_{j=1}^{J} PRODUCT\ USAGE_{ij,t} * CONTRIBUTION\ MARGIN_{i,t}$$

Where:

$Profit_{i,t}$ = current and future (predicted) contribution margins from the customers of the company (i) each time period (t, $1 \leq t \leq T$),

j = index for banking products ($1 \leq j \leq J$, J is the total number of products),

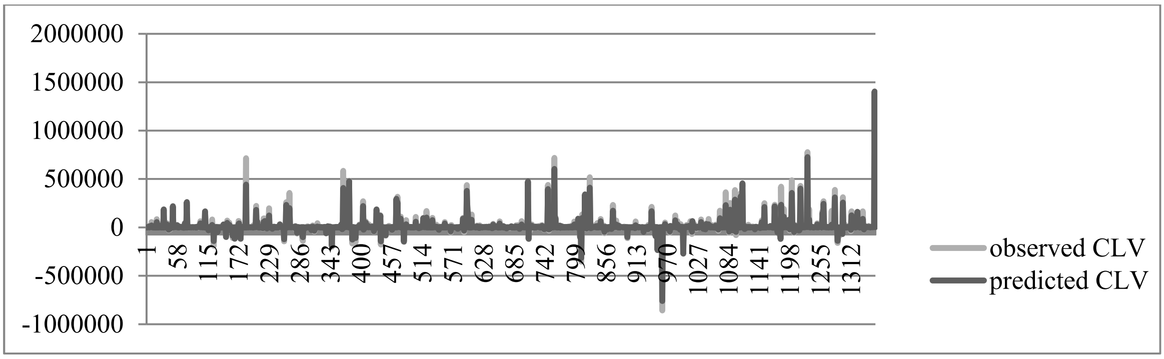
$PRODUCT\ USAGE_{ij,t}$ = observed and predicted values of the second component of CLV, and

$CONTRIBUTION\ MARGIN_{i,t}$ = observed and predicted values of the third component of CLV.

Secondly, we calculate the $\boldsymbol{observed\ CLV_i}$ using the observed values about product usage and contribution margin. We also calculate the observed CLV using the same formulas that we have indicated above.

Therefore, proceeding in the same way that with product ownership, product usage and contribution margin, Figure 27 shows the comparison between the $\boldsymbol{observed\ CLV_i}$ (the light grey line) and the $\boldsymbol{predicted\ CLV_i}$ (the dark grey line) for all the customers in the sample.

Figure 27. Comparison between observed and predicted CLV



As one can appreciate, for most of the customers, the proposed Bayesian hierarchical models seems to produce a good fit over the observed results of the drivers of CLV. Once more, this is because the differences among the two lines are small and the distribution of the data is quite similar. In general, it seems that our models perform well to estimate CLV. Additionally, Pearson correlation (R) is equal to 0,977 ($p < 0,01$), which implies that a very strong positive relationship exists between predicted CLV and observed CLV.

6.4.5. Value-based segmentation

The R software version 3.0.2 is selected to perform the regression tree analysis using the tree package. In order to validate the results obtained, we split the sample in two sub-samples, analysis or training sample (1.000 customers) and hold-out or test sample (357 customers).

Firstly, using the analysis sample we get a tree with twelve nodes. This first tree is pruned in order to get a better interpretation of each group and avoid those groups with a small number of customers. Finally, after pruning the tree we get five nodes or customer segments. Table 20 shows the node number, the number of cases in each node, the node deviance, the average value of predicted CLV for each node and the 95% C.I. of the five segments of customers obtained. Figure 28 also shows a graphical representation of the tree indicating the variable used to split, the split criterion used in each division of cases and the average value of the predicted CLV for each node. As you can see in Table 20 and Figure 28, gender variable was automatically excluded of the analysis by the algorithm because this variable does not offer possibilities to split the sample. Therefore, we explain each of the obtained nodes (i.e., segments) according to their average value of predicted CLV, average income and age of the customers of each segment, as follows:

- The first node is formed by 728 customers (72,8% of the analysis sample) whose income < 14.581,2 euros, and whose average value of predicted CLV is equal to 2.518,5. According to the value of the predicted CLV, this is the fourth segment of customers (in descending order, that is, arranged from largest CLV to smallest CLV

- The second node contains 156 customers (15,6% of the analysis sample). These customers are characterised by income ∈ (14.581,2, 47.038,5) euros and age < 53,5 years. According to the value of the predicted CLV (equal to 38.113,59), this is the third segment of customers (in descending order).

- The third node contains 23 customers (2,3% of the analysis sample). These customers are characterised by income > 47.038,5 euros and age < 48 years. According to the value of the predicted CLV (equal to 56.026,21), this is the second segment of customers (in descending order).

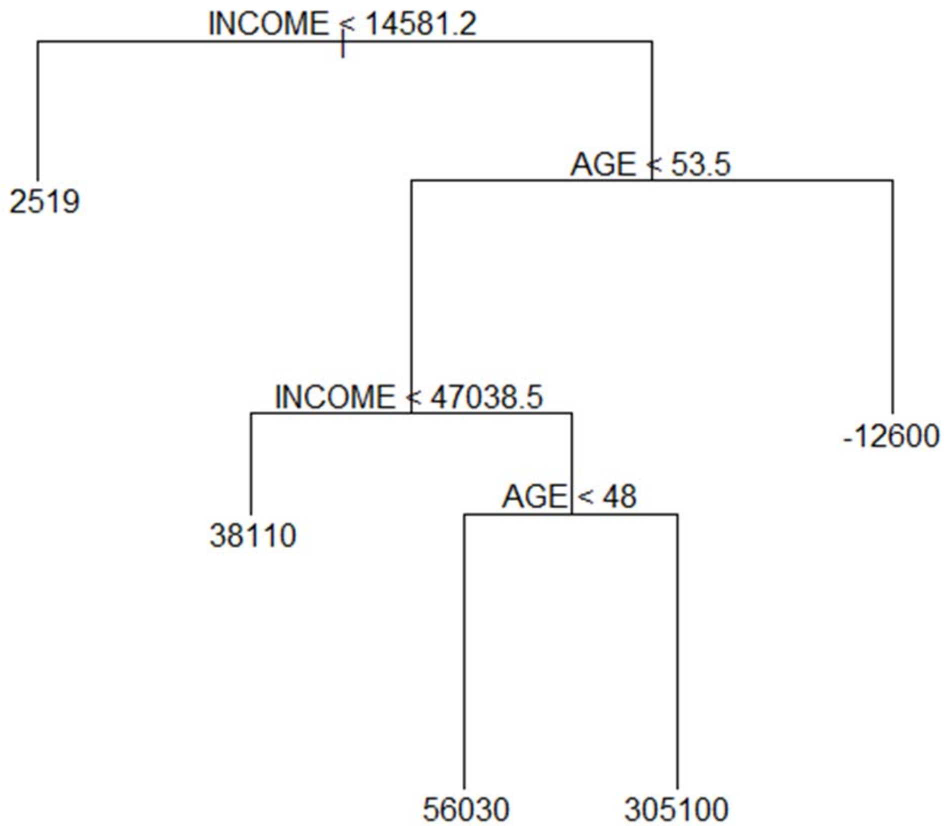
- The fourth node contains 5 customers (0,5% of the analysis sample). These customers are characterised by income > 47.038,5 euros and age ∈ (48, 53,5) years. According to the value of the predicted CLV (equal to 305.100), this is the first segment of customers (in descending order), or in other words, this is the most valuable segment of customers. Due to the low number of cases in this node, we have to be cautious with its interpretation. Maybe it is composed mainly by outliers. Therefore, in order to know the profile of the most profitable customers, we can conclude with a combination of the characteristics of the segments number 3 and 4 (the most valuable ones): they are customers mainly characterised by income > 47.038,5 euros and age < 53,5 years.

- The fifth node contains 88 customers (8,8% of the analysis sample). These customers are characterised by income > 14.581,2 euros and age > 53,5 years. According to the value of the predicted CLV (equal to -12.596,53), this is the fifth segment of customers (in descending order), or in other words, this is the least valuable segment of customers.

Table 20. Regression tree from the value based *ex poste* segmentation proposed

Node number	Number of cases	Node deviance	Value of predicted CLV	95% C.I.		order of the segments (according to the predicted CLV)
				lower limit	upper limit	
1	728	1,47E+14	2.518,85	1.484,38	3.553,32	4
2	156	1,44E+15	38.113,59	22.858,49	53.348,68	3
3	23	5,24E+14	56.026,21	-10.694,08	122.746,49	2
4	5	3,63E+14	305.115,99	-69.141,66	679,373,63	1
5	88	5,47E+14	-12.596,53	-29.391,71	4.198,65	5

Figure 28. Regression tree from the value based *ex poste* segmentation proposed (*)



(*) In each terminal node you can see the average predicted CLV value for this node

Secondly, using the hold-out sample we validate the results previously shown. For this task, we use the multi-group discriminant analysis in order to find a linear combination of variables (in this case, age, gender and income), which characterises the two samples that we use to define and validate the tree, i.e., analysis and hold-out sample, respectively. From the tree with 5 nodes, we delete those groups or nodes with a small number of cases, such is the case of node number 4 (with 5 cases). Therefore, we work with only 4 groups and 352 cases (357 – 5 cases from node 4). Additionally, we identify and delete one outlier in the hold-out sample (case number 245), mainly because its predicted CLV is equal to -760.664,42 (the lowest value in the sample). Therefore, the final sample size for the hold-out cample is 351 cases. We use a variable that contains the information regarding the segment where each individual is located as the grouping variable and age, gender and income as discriminators to perform the previously mentioned multi-group discriminant analysis. Every discriminant analysis is followed by an ANOVA test for validation of the results of the discriminant analysis. Some important calculation details are shown in Table 21.

Table 21. Discriminant analysis for the validation of the regression tree

Linear method for response: automatic classification				
Predictors:	age; gender; income			
Group:	*1*	*2*	*3*	*4*
Count:	248	55	7	41

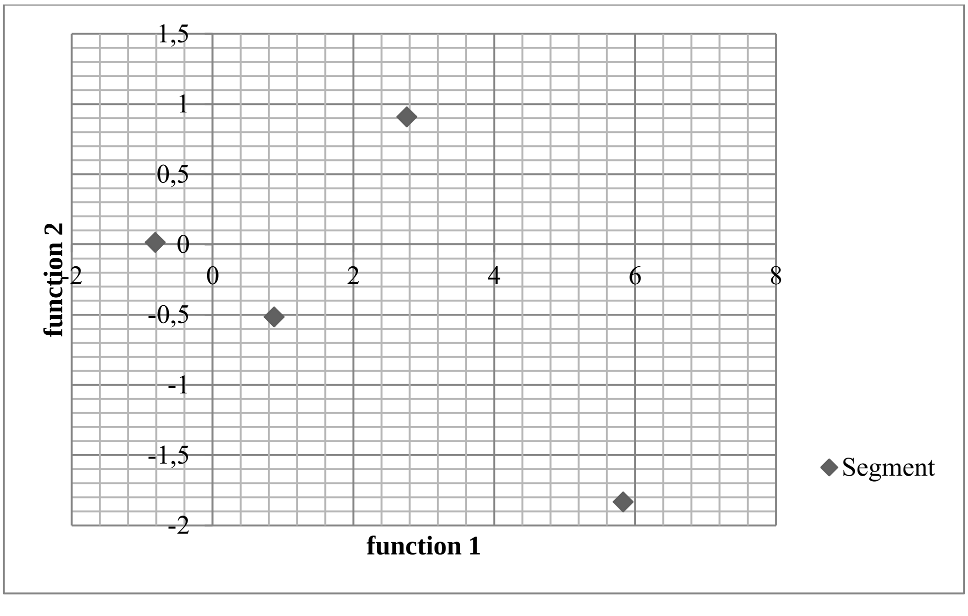
Summary of the classification				
	True group			
Put into group	*1*	*2*	*3*	*4*
1	248	0	0	0
2	34	16	0	5
3	0	0	6	1
4	13	1	8	19
Total n:	*295*	*17*	*14*	*25*
n correct:	*248*	*16*	*6*	*19*
Proportion:	*0,84*	*0,94*	*0,43*	*0,76*
n = 351; n correct = 289				
Proportion correct (hit rate) = 0,82; Proportional chance criterion = 0,61; z = 8,05 (p < 0,01)				

ANOVA for the automatic classification					
	Wilks' lambda	*F*	*gl1*	*gl2*	*p value*
age	0,762307676	36,06559358	3	347	0,00
gender	0,970169225	3,556520072	3	347	0,02
income	0,320502883	245,2245223	3	347	0,00

The measure of predictive accuracy used is the *hit ratio* or the percentage of cases correctly classified (Hair *et al.*, 2009 p.266), that is equal to 0,82. Moreover, comparisons between the hit ratio and the proportional chance criterion (Hair *et al.*, 2009 p.365) are also made. Particularly, this proportional chance criterion is equal to 0,61 (z = 8,05, p < 0,01). Therefore, our results are statistically significant demonstrating that the classification obtained using the analysis sample is also applicable in the hold-out sample.

From the three canonical discriminant funtions calculated, two are significant (p < 0,01). Accordingly, the first one is mainly determined by income ($r_{income,canonical\ function\ 1} = 0{,}985$) and explains 91,3% of the difference between the four groups. The second canonical discriminant funtion is mainly determined by age ($r_{age,canonical\ function\ 2} = 0{,}895$) and explains 8,7% of the difference between groups. Finally, the third canonical discriminant funtion is mainly determined by gender ($r_{gender,canonical\ function\ 3} = 0{,}927$), but this function does not explain any difference between the four groups (as we have explained before, this variable was also automatically excluded in the regression tree analysis). Finally, we enclose Figure 29, which represents a two dimensional graph with the centroids of the four segments defined by using the only two significant canonical discriminant functions. Canonical discriminant function number 1 is represented in the *x* axis, whereas canonical discriminant function number 2 is represented in *y* axis.

Figure 29. Centroids of the significant canonical discriminant functions



To sum up, according to the results obtained we can conclude that for a sample of customers that are less than 53,5 years old, when their income is higher and their age is also higher, their value (in terms of CLV) is also higher. On the contrary, for a sample of customers older than 53,5 years, when their age is higher, their value (in terms of CLV) is lower. Particularly, the most valuable customer group for the bank is composed by those customers characterised by income > 47.038,5 euros and age < 53,5 years (segments 4 and 3), whereas the least valuable customer group is composed by those customers with income > 14.581,2 euros and age > 53,5 years (segment 5).

6.4.6. Customer Equity

Finally, we can also calculate CE. Thus, following the suggestions of Rust *et al*. (2000, 2004a), we adapt their formula to our context and data available. CE is specified as follows:

$$CE = mean(CLV_i) * POP$$

Where:

> *mean* (CLV_i) = the average lifetime value for firm customers (i) across the sample, that is 9.455,66,
>
> *POP* = the total number of customers in the sample, that is 1.357 customers.

Therefore, CE is equal to: 12.831.330,62. This quantity is an over-estimation of the real value (in euros term) of the customer base of the collaborating bank. This is because we have not calculated CLV taking into account the contribution margin related to each product because it was an unavailable measure. Due to the data available by the bank, the result of the CLV expression is only a proxy variable of the amount that is assumed to be in Euros. Contribution, however, is not expressed as an amount in Euros per unit of the product and is given for each customer by the bank. This aspect is going to be considered in future research streams (for more details see limitations and future research streams sections in Chapter 7).

Chapter 7. CONCLUSIONS, MANAGEMENT IMPLICATIONS, LIMITATIONS AND FUTURE RESEARCH STREAMS

7.1. Conclusions

Customer value, considered as an important firm asset, has been measured and assessed through different techniques. Simple revenue or profit has often been chosen as a measure of customer value, however in an increasingly competitive world it has been recognised that customers whose revenues per period are lower, but whose loyalty is greater, may be better customers during a longer time horizon (Drew *et al*., 2001). This is the main argument that justifies our selected framework based on Customer Lifetime Value (CLV). This measure assesses the value from customers taking into account several components and drivers of CLV, providing important diagnostics about the future health of a business, which may not be obvious from traditional financial metrics.

Within this research the importance and validity of CLV and its aggregation (i.e., CE) have been highlighted to assess the customer-base and, by extension, the firm. For this reason, and given the great number of CLV and CE models developed until now, in this research a classification of a set of published researches about CLV and CE models have been performed according with several criteria, such as *type of relationship between customer and company*, *if the analysis is historical or predictive*, *deterministic or stochastic*, *source of data*, *if the effect of competition is included* and *level of aggregation in the CLV calculation*. This classification serves as a guide with key requirements for developing these types of models and it has helped us to establish the main characteristics of our own model. Particularly, we have chosen a *contractual setting, predictive and stochastic analysis (probability and data mining models), with data from the company database, without competition effect (this information is not available for this research)* and finally, according to *the level of aggregation in the CLV calculation we get individual CLV measures as an input for an ex poste segmentation of customers*.

Despite many efforts from researchers to drive the implementation of customer value management and the related models –for instance through churn tournaments (Neslin *et al*., 2006) or implementing NBD-models in Excel to facilitate their usage (Fader *et al*., 2005b)– practitioners are still reluctant to adopt the suggested models, mainly because although there are many models for this purpose, most of them are theoretic, complex and not applicable. To alter this, researchers have to clearly demonstrate and communicate that their models outperform the heuristics typically used by practitioners. In addition, researchers have to continue their efforts to make their work more accessible, by for instance implementing their models in standard software. On the other hand, it is also desirable that more marketing executives consider implementing state-of-the-art-models (Verhoef *et al*., 2007). For the purposes of this research we have developed a model that covers an important number of products and predictors of CLV in an extremely complex context. In this case we had to look for simplicity inside the inherent complexity of the problem in order to build a model easy to use by the bank. Therefore, Bayesian statistic was the key to solve our CLV problem (Ntzoufras, 2009), in particular using Hierarchical Bayesian models. We have also used data mining methods, in particular regression trees, to perform an *ex poste* segmentation of the customer base. In this way we have demonstrated that company databases, which provide time-series data on individual customers, are a valuable tool for explaining buying behaviour and it can potentially lead to increase sales.

The results obtained in this study demonstrate the usefulness of the proposed Hierarchical Bayes model in the analysis of consumer-information extracted from a company database. In particular, the methodology developed has two main strengths: (1) it is possible to model together different issues (with different statistical distributions), combining product ownership, product usage and contribution margin, using panel transaction data and consumer personal data, and (2) it is possible to estimate the separate effect of consumer characteristics (i.e., independent variables) on product ownership, product usage and contribution margin. In our case, some of the covariates used in the analyses were found to be good predictors for explaining product usage and contribution margin (not in case of product ownership). In particular:

(i) Estimation results about the **product ownership** models have been quite poor, implying that the predictions obtained from these models are inaccurate (as you can see analytically in Tables 18 and graphically in Figures 24, both in Appendix D). Therefore, we have decided to discard the product ownership term from our CLV model formula. The estimation of CLV does not suffer any modification in global terms, because product usage contains the same information as product ownership but in a more detailed way (product ownership is a set of 18 binary variables and product usage is a set of 18 categorical variables).

(ii) Among the covariates that we have taken into account in order to predict **product usage** for each of the product considered, one-period lagged of product usage (with a positive influence) has been the variable that has exerted the most important effect over the usage of the different products, except for account and credit. The driver average monthly liabilities has been the variable that has exerted the most important effect over the usage of account (with a positive influence) and average monthly liabilities in case of the usage of credit (with a negative influence). Additionally, in the particular cases of credit card, debit card, home insurance, home loan and investment fund, those customers who have adopted online banking, use more products than the traditional customer population. Therefore, adoption of online banking influences the product usage (for the previously mentioned products), as was stated by Hitt and Frei (2002).

(iii) Among the covariates that we have taken into account in order to predict **contribution margin**, total quantity of purchases (with a positive influence over contribution margin), cross-buying (with a negative influence over contribution margin), one-period lagged of contribution margin (with a positive influence over contribution margin) seem to have the largest impact on this contribution margin. Other important predictors are: average monthly assets (with a positive influence over contribution margin) and average monthly liabilities (with a negative influence over contribution margin). Adoption of online banking does not have a significant effect over the contribution margin. Therefore, regarding the result that emerges from Hitt and Frei (2002) (i.e., PC banking customers offer a higher contribution margin than the traditional customer population), we cannot give support to this idea.

Additionally, the proposed Bayesian hierarchical models seem to produce a good fit for most of the customers over the observed results related to product usage and contribution margin (not in case of product ownership). This fact is observed firstly analytically in Tables 19 (related to product usage) and using Pearson correlation (related to contribution margin), and secondly graphically in Figures 25 (related to product usage), Figure 26 (related to contribution margin) and finally, Figure 27 (related to CLV). In all these Figures the differences between the two lines are small and the distribution of the data is quite similar. In general, it seems that our models perform well to predict product usage, contribution margin and therefore, CLV.

Our analyses also reveal that data from a company database are useful not only for identifying the most/less valuable individual customers, also to segment them. Using the results obtained in the segmentation proposed banks are able to know which are those more profitable customers from measures such as CLV_i, age_i, $gender_i$ and $income_i$, where *i* refers to each customer. In Figure 28 you can see more details about this segmentation and the different groups obtained. Particularly, according to the results obtained we can conclude that for a sample of customers that are less than 53,5 years old, when their income is higher and their age is also higher, their value (in terms of CLV) is also higher. On the contrary, for a sample of customers older than 53,5 years, when their age is higher, their value (in terms of CLV) is lower. More specifically, the most valuable customer group for the bank is composed by those customers characterised by income > 47.038,5 euros and age < 53,5 years, whereas the least valuable customer group is composed by those customers with income > 14.581,2 euros and age > 53,5 years.

Therefore, using this *ex poste* segmentation we have answered one of the research questions described in Chapter 1, in particular: Can we rank and order the customers of the bank according to their value? Indeed, we have not only ordered customers according to their value, we have got different groups of customers according with their value and several socio-demographic variables of each one.

We have also calculated the overall value of the bank customer base. To perform this task we have used the CE concept that has allowed us to obtain this overall assessment of the customer base. This is an important measure that could help to make managerial decisions because, as we have noted along this research, customers are considered the main asset of the company. However, CE in our case is an over-estimation of the real value (in euros term) of the customer base of the collaborating bank. Due to the data available by the bank, the result of the CLV expression is only a proxy variable of the amount that is assumed to be in Euros. Contribution margin variable is not expressed as an amount in Euros per unit of the product and is given for each customer by the bank. This aspect is going to be considered in future research streams.

According to the last research questions described in Chapter 1 (i.e., How could the bank improve CRM?), we have also asked to this question. Understanding how to drive CLV and/or CE is central to the decision making of any firm and formulating a procedure to achieve this objective can give the firm an important competitive advantage (Aravindakshan *et al*., 2004). Predictions about the CLV, such as the output of our model, are an important input to target customers for special treatment, which is a central operational tactic of relationship management (Drew *et al*., 2001). More valuable customers should be treated in special ways in order to enhance their profit production and increase the profitability of them being retained. On the other hand, less valuable customers should be offered a product or service that is less costly to provide.

Additionally, in this research it has been posited that customer valuation is mainly based on the principles of contemporary finance of assets' valuation, more precisely the discounted cash flow (DCF) method. CLV (and by extension CE) has been differentiated from CP and DCF, but the main idea that emerges from these related techniques is that CLV comes from these financial measures (that is, DCF and CP). According to the financial origin of CLV-CE, two important aspects that should be considered by researchers are: (i) *how to calculate the monetary value that each customer brings to the firm* and (ii) *how to calculate the present value of this monetary value*. According to (i) the first idea, some researchers argue that CLV is based on the difference between customer revenues and customer costs (e.g., Calciu and Salerno, 2002; Gurau and Ranchhod, 2002; Mulhern, 1999), while other researchers propose contribution margin as this monetary value (e.g., Berger and Nasr, 1998; Malthouse and Blattberg, 2005; Reinartz and Kumar, 2000). Nevertheless, according to the financial theory, the value of any asset is the present value of its cash flows over time (i.e., cash inflows minus cash outflows), issue that few researchers have accurately applied in their CLV models (an exception is Buhl and Heinrich's (2008) research). According to (ii) the second idea, it is also needed a discount rate to estimate CLV used to transform expected future cash flows into a present value. The discount rate has to reflect the riskiness of the cash flows (Damodaran, 2002). Some researchers argue that discount rate is based on the lending rate that is appropriate for the time of the study (e.g., Venkatesan and Kumar, 2004), or depends on the general rate of interest and is normally proportional to the treasury bill or the interest that banks pay on saving accounts (Kumar, 2008a p. 48). Nevertheless, according to the financial theory, the Weighted Average Cost of Capital (WACC) is the method used to discount customer cash flows (Ryals and Knox, 2007). Therefore, two important suggestions or future research streams emerge from these two arguments: (i) the present value of future cash flows over time is the most suitable way to measure the monetary value that each customer brings to the firm, and (ii) WACC is also the most appropriate method to get the discount rate.

According to the theoretical influences of CLV-CE, it has been noted that Srivastava el al.'s (1998) application of RBV theory to marketing management identified a particular type of resource: the *market based asset*. This allowed customers and their relationships with the firm to be treated as critical resources that fitted the RBV criteria of value, i.e., rarity, inimitability and non-substitutability and that should be developed, augmented, leveraged and valued in a similar way to the traditional resources of the firm. Clearly, the calculation of CLV-CE is aligned with market assets based perspective because it recognises the worth of customers and customer relationships as assets of the firm. In fact, CLV-CE perspective views customer assets as *super-assets* (Hogan *et al*., 2002 p. 7) being superior to all other resources and assets. The value of these super-assets is determined by the choices the firm makes to combine and apply its other resources in the market. Following Osborne and Ballantyne (2012), this perspective is limited as it establishes the value of customers *to* the firm, or in other words, it answers to the question: "*How the firm captures value from its resource bundle?*". However, it does not provide any explanation about how value is created and customers are not seen as active participants in the value creation process. On the contrary, customers are passive receivers of predetermined value *for* the firm, providing an input to the own value creation process of the firm.

Trying to get a better understanding about the review about CLV-CE postulated by Osborne and Ballantyne (2012), it is interesting to formulate their same question: "*Can customers as super-assets support a claim of customer centricity?*" Within this research it has also been posited that other previous authors viewed CLV-CE as the most appropriate perspective to build a customer-centric organization (Bell *et al*., 2002 p. 78; Hogan *et al*., 2002 p. 4; Jain and Singh, 2002 p. 35; Rust *et al*., 2004a p.110; Verhoef and Lemon, 2013 p. 5). However, despite the CLV-CE claim of customer centricity, Osborne and Ballantyne have noted that this framework examines a value through the eyes of the firm and the assessment of this value is calculated by the worth of customers *to* the firm. Therefore, this perspective still considers customers as passive participants at the end of the value creation process of the firm, provides no insight into how value for the customer is created and the focus is the efficiency of the activities of the firm. Thus, following Osborne and Ballantyne's (2012) claim we can conclude recognising that within our selected perspective the marketing system is still one-sided and firm centric. In section 7.4 about future research streams we propose different ways to deal with CLV from a wider perspective, taking into account measures directly from customers, such as the voice of customers.

7.2. Management implications

Customer Value Management (CVM) has its roots in relationship marketing (Verhoef and Lemon, 2013). Its core goal is to determine and maximise the value of a company customer base analysing individual data on prospects and customers. In particular, CVM takes into account customer-centric measures, such as CLV. As a consequence, marketing becomes more accountable, which causes less waste of marketing spending and more effective allocation of marketing resources over customers and marketing instruments.

CVM can improve business performance because it allows acknowledging that customers differ in value and firms can act on these differences (Verhoef *et al*., 2007). This improvement in business performance is spread in three ways: (1) CVM is a market-based resource for competitive advantage, (2) CVM increases the customer-centric orientation of the firm, and (3) CVM leads to more accountable marketing (Verhoef and Lemon, 2013 p. 2).

In this research, as we have noted previously, we have developed a model of CLV and a subsequent *ex poste* value-based segmentation as a way to improve CVM strategies. In particular, this research highlights the importance of CLV in CVM providing a framework to assess customer base in a Spanish financial service company. This is a clear implementation of the core of CVM but it requires additional efforts to achieve the desired improvement in business performance. Verhoef and Lemon (2013) have suggested several important lessons that firms can employ to get a successful CVM and that should be considered by any firm wishing to implement CVM. Such lessons are summarised below and they are also key for a succeed implementation of CVM in our selected context:

(i) Ensure that CVM is more customer driven than information technology driven. Technology investment in CVM (e.g., software, hardware) should be driven in such a way that benefits customer-centric processes within the organization. Applying more technology not always is the correct way to solve problems and firms need to plan a customer strategy before implementation.

(ii) Invest in strong analytical capabilities as the process of extensively employing data, quantitative analyses, statistical models and fact-based management techniques to drive firm decisions and actions. Analyses may encompass disparate fields such as identifying potential customers, predicting response behaviour of existing customers, calculating the costs of maintaining a relationship and cross-selling predictions.

(iii) Understand the key drivers of customer acquisition, retention and expansion. Adopting a customer-centric view by actively measuring and monitoring customer metrics is not sufficient. To truly succeed in managing customers for maximum value, firms must clearly understand what drives customer acquisition, customer retention and customer growth over time measuring and analysing customers' perceptions (e.g., value equity, brand equity and relationship equity, for more details about these measures see Lemon *et al*., 2001).

(iv) Manage channels to create customer value. Firms frequently assume that providing more channels can create a stronger customer experience, but more channels sometimes imply more complexity. A multichannel strategy requires a strong analysis of the consequences of adding channels, migrating customers to other channels and eliminating channels. According with our results related to the product usage models, the predictor variable adoption of online banking exerts an important influence over the usage of credit card, debit card, home insurance, home loan and investment fund. We can conclude that the maturity of this kind of customers (because of the type of products that they use) leads them to use online banking. In other cases, such as stock capital, saving insurance, (not) linked life insurance, other insurances, account, deposit, pension plan, securities, (micro) consumer loan, mortgage and credit, adoption of online banking do not exert any influence over the usage of this product, maybe because this kind of customers are more conservative and sometimes their age is higher (especially, for example, in case of pension plan). Therefore, these ideas could help the bank to manage the online channel according with the type of customer considered.

(v) Managing customer engagement. Firms must create committed customers as one of the strongest tools to ensure customer retention. Social media and other new media help companies to develop non-transactional behaviours (e.g., word of mouth, blogging, customer ratings) to strengthen customer engagement, although firms should develop a set of capabilities to manage it.

(vi) Managing customer networks. These networks are a source for customers that search information, buy products and communicate. Understanding these customer networks (e.g., identifying customers with high social influence) will become increasingly important to managing customer value.

(vii) Managing the customer experience. Creating superior customer experiences, which can be shaped by sensory, affective, intellectual and behavioural experiences, seems to be one of the central objectives to foster customer loyalty and thus, it has also an important role in CVM.

7.3. Limitations

The idea that customers are important firm assets has led to the development of a large number of methods for estimating CLV. These methods has been considered as an important strategic marketing tool, because they have helped firms to quantify customer relationships (Berger and Nasr, 1998), have illustrated the profitability of customers (Reinartz and Kumar, 2000; Wiesel *et al.*, 2008) and have provided references for the allocation of marketing resources to customers and market segments in order to maximise CLV (Blattberg and Deighton, 1996; Kumar *et al.*, 2008b). However, existing CLV models still have limits in applicability for three reasons (Wang and Hong, 2006):

(1) Some of the CLV models developed until now, which predict purchase behaviour based on past customer spending patterns or demographics characteristics, are of limited use in predicting future behaviour (Libai *et al.*, 2002). On one hand, additional factors must be considered, as social effects, competitive effects, economic environment, product lifecycle, customer lifecycle, customers' purchasing habits, lifestyle, customer satisfaction, price sensitivity and brand loyalty (Hogan *et al.*, 2002; Jacobs *et al.*, 2001; Mulhern, 1999; Stahl *et al.*, 2003). These factors are going to allow an extension of the CLV basic model and effectively apply it to a complicated open market. On the other hand, the probability-based CLV models only guarantee that the models predict well within the time horizon of the collected data, but there is no guarantee for forecasting values beyond that horizon (Bell *et al.*, 2002).

(2) Existing CLV models provide a static estimate of customer value for a given future period. Using CLV customers could be segmented into several levels of the customer pyramid of the firm, such as profitable, less profitable and unprofitable (Zeithaml *et al.*, 2001). However, dynamic markets require a more tactical view towards these measures for segmenting customer base. For example, the *direction of customer profitability* is a reliable indicator of the customer's status (defecting, upgrading or steady) and *volatility of customer profitability* represents the possible risk level of a customer's profitability for a firm (relatively unsteady customers or relatively steady customers). Therefore, a mechanism to monitor both indicators (if data are available) would enable firms to

dynamically adjust marketing activity towards their targeted customers (Wang and Hong, 2006).

(3) Further information of customer accessibility, needs and customer attitudes (e.g., preferences and satisfaction) is needed to incorporate customer profitability measures into marketing planning (Jain and Singh, 2002). Analysis of customer profitability is often used to indicate possible consumption patterns of the targeted customers; however it is not enough for identifying the customers a firm truly wishes to acquire or retain through allocating additional marketing resources.

Based on the previously shown general limitations of most CLV models we can conclude admitting some limitations of our model, firstly, regarding measures and the selected sample, and secondly, about the selected methodology to solve the problem.

Regarding measures of the variables used for this research, we have not measured monetary value or contribution margin of customers for each of the different types of products considered. As we have noted previously, for this reason the calculated CE is an over-estimation of the real value (in euros term) of the customer base of the collaborating bank. Also regarding monetary value of customers, it has not measured as the difference between customer inflows and outflows (cash flows) neither we have used WACC (from finance) to calculate the discount rate. We have also measured a limited array of behavioural and socio-demographic variables, not considering: (1) important information to get CLV derived from individual transactions of customers, social effect, economic environment, competitive effect, neither measures of customer perceptions, attitudes or in general, V.O.C. (e.g., customer satisfaction and customer preferences); (2) more socio-demographic variables that can enrich the second empirical stage offering more information in order to get a more accurate profiles of customers. Additionally, as we have noted in Chapter 6, the choice of the sample and the subsequent two years as the observation period available has certain limitations. Firstly, regarding the use of data from a single sample, this gives rise to sampling error, i.e., the inaccuracy of results that occurs when a population sample is used to explain the behaviour of the total population (Kumar *et al*., 2006b). Additionally, regarding the subsequent two years as the observation period, this fact has an impact on age of customers (generating bias towards young families), but also on length of the relationship and the opportunity for cross-selling.

The traditional approach to estimate probabilities is vulnerable to sampling error because of the implicit assumption in all regression analyses that the coefficients of the independent variables of the sample group are representative of the population as a whole. If the sampling error is severe enough, the company using this methodology can end up choosing the wrong product to push at the wrong time to the wrong customer and even using the wrong channels (channels that a company uses are often a big determinant of both product choice and purchase timing). Unfortunately, most companies have no option and often have to rely on relatively small samples to perform the calculations. They frequently lack enough data on all their customers to estimate meaningful relationships between the drivers of purchasing behaviour. So how can companies derive probabilities free of sampling error? The answer lies in a branch of statistical mathematics called Bayesian estimation. The methodology has been around for decades but is only recently entering the marketing mainstream. Therefore, despite the previously mentioned limitations (regarding measures and sample), Bayesian estimation overcomes some of the problems related to regression models (for more details about these limitations see section 4.1.1.1). Rather than estimating a single coefficient for each variable (as regression analysis does), the formula at the heart of this technique first specifies the range of coefficients that could have produced the observed data of the sample being analysed. Then, through an iterative chain of calculations, it allows the analyst to determine the most probable coefficients for the variables involved, those that would most likely have produced the observed data. You can think of Bayesian estimation as reproducing the dots on a scatter diagram rather than finding the best-fit line, which is what regression analysis does. This kind of calculation has greater predictive power because it reproduces the actual behaviour of a sample rather than estimating a set of coefficients from one sample and then assuming that those coefficients are valid for the whole population. However, mainly from a methodological point of view we can improve our hierarchical Bayesian model testing other distributions for product ownership, product usage and contribution margin (for example, zero-inflated Poisson for product usage), and also searching more informative prior distributions for the parameters of the model. Additionally, we have predicted the values of the dependent variables for only one year, mainly because a hierarchical Bayesian model needs the values of the independent variables in order to predict the values of the dependent ones. However, the longer the span of period over which the data are collected the better is (Kumar, 2008b p. 81) and the goal should be to work with a data period that is broad enough to reflect the reality of the marketplace. Furthermore, it could be interesting to find new methodologies that help us to predict the values of the dependent variables without knowing the values of the independent ones, as well as to work with customers with different starting points of their relationship (left-censored data). Finally, it could be interesting to improve the second methodological stage using different segmentation methods and get a comparison between them.

After having reviewed the most important limitations of this study, we propose some future research streams to overcome these drawbacks in the following section.

7.4. Future research streams

In this research it has been developed a model to estimate CLV and segment the customer base combining hierarchical Bayesian analysis (probability model) and regression tree analysis (datamining model). This combination was formulated because despite the fact that the assessment of customer is an important trend in various disciplines such as accounting, finance and especially in marketing, multidisciplinary approach is needed to complement the models developed to date. The challenge is to establish a dialogue between marketing and finance (Bauer and Hammerschmidt, 2005; Wiesel *et al*., 2008), as well as dialogue between marketing and the discipline of computer science (Gupta *et al*., 2006; Rust and Chung, 2006) with the objective of integration between different ways of modelling and the marketing measures. In particular, although we find in finance the origins of CLV and it is an important support to calculate CLV-CE, continuous advances in information and communication technology have also had an important role in the development of this framework. They have allowed companies to collect large amounts of customer data at a reduced cost and consequently, these companies have been forced to acquire skills to store, share, analyse and transfer valuable information from these data. The objective is to guide marketing strategies and gain control (direct, optimise and automate) over the decisions they make every day (Apte *et al*., 2003). To aid companies in these tasks, computer science discipline brings advanced (also known as predictive) analytics techniques that combine information on past circumstances, present events and projected future actions to answer questions or solve problems (Bose, 2009). These techniques are applied to get an automated extraction of ‘hidden’ predictive information from databases, especially in companies with a strong customer focus. In particular, advanced analytics are classified into several groups: data processing, prediction, regression, classification, clustering, link analysis (associations), model visualization and exploratory data analysis. Examples of data mining methods are: statistical methods, case-based reasoning, neural networks, decision trees, rule induction, Bayesian belief networks, genetic algorithms/evolutionary programming, fuzzy sets and rough sets. They are used in combination with one another to gain information, analyse information and predict outcomes of the problem solutions (e.g., in the areas of sales forecasting, direct marketing, customer acquisition, retention and extension purposes and marketing campaign analysis). Therefore, managers also can use advanced analytics with data mining to model CLV-CE and to get more accurate analysis. Applying this idea to this research, as a future research stream it can be interesting to prove different methodologies from the discipline of computer science (suitable for our context and data available) to test whether our results are improved. For example, regarding the first empirical stage, it could be interesting to find new methodologies that help us to predict the values of the dependent variables without knowing the values of the independent ones. Additionally, we have used the data for the first 12 months for analysis, while the data for the last 12 months as a holdout sample for validation. We propose to apply another type of validation consisting of comparing the estimations that result from different methods (method validation). For example, we can compare Bayesian estimation results with those that were found using a (multivariate) logit or probit analysis. This alternative applies only when such results are available (this is the case of product ownership, following the example of (multivariate) logit or probit analysis). Regarding the second methodological stage, it can be interesting to use different segmentation methods and also get a comparison between them as another way to validate results from regression tree.

Within the Bayesian framework, and as we have previously noted, we have predicted the values of the dependent variables for only one year, mainly because a hierarchical Bayesian model needs the values of the independent variables in order to predict the values of the dependent ones. In this case, we remark as a future research stream to work with a wider observed period of data in order to predict more periods for the dependent variables. We can improve the estimation using Bayesian statistics testing other distributions (with better accuracy) for product ownership, product usage and contribution margin (for example, zero-inflated Poisson regression in case of product usage because of the large amount of zeros), and also searching more informative prior distributions for the parameters of the model. There is also the possibility to use a super computer to estimate the complete model (all the 37 equations at once) including possible interaction effects that may exist between the bank products.

Other possibility for further research is to take into account the two suggestions that emerge from finance, they are (i) the present value of future cash flows over time is the most suitable way to measure the monetary value that each customer brings to the firm and (ii) WACC is also the most appropriate method to get the discount rate. Therefore, we propose to measure monetary value of customers as the difference between customer inflows and outflows (e.g., Buhl and Heinrich, 2008) and calculate discount rate using W.A.C.C. (e.g., Ryals and Knox, 2007).

It also would be interesting to consider more variables to estimate the CLV (Woo *et al*., 2005), for example:

- Contribution margin measured for each type of product. The suggestion then is to find an estimate of the contribution per product. A linear regression of contribution margin as a function of product usage (for 18 products) will lead to an estimated β_j whose value can be interpreted as an estimate of the contribution of product j per unit of time. The estimates of per unit contributions can be used in future research.

- Transaction information from own customers (such as the amount of money that customers spend using debit or credit cards). This information is actionable for own customers of the financial services company, but currently it is not available for competitor's customers or latent customers.

- Effect of competition. It is important to point out that most modelling approaches ignore competition because of the lack of competitive data. Understanding what drives customers to the competition is also critical because it can help companies to answer why customers do not buy from their company, which can be very informative (Verhoef and Lemon, 2013). If the model explicitly considers the relationship between the focal brand and the competitor's brands, it will allow the creation of models that contain both customer attraction and retention in the context of brand switching. The main advantage is that competitive effects can be modelled, thereby yielding a more accurate account of CLV and CE (Aravindakshan *et al*., 2004).

- Lifestyle or preference data also will be useful. The lifestyle, preference, attitudinal data could be induced from customer surveys and extended enterprise behaviour (Bloch and Pigneur, 1995).

- VOC. With the term voice of customer (VOC) is known all kinds of communication messages from customers to the bank through customer contact channels (Woo *et al*., 2005). It includes asking, claiming on public mediation institutes, complaining, commending or praising. In the company, VOC means mainly customer complaints that are collected through call centres, Internet homepages, ARS systems, etc. Managing customers' dissatisfaction leads to identification of customers' unsatisfied requirements. VOC is recorded and maintained under the VOC code structure according to VOC characteristics and it is also configured as interesting information to consider for further research.

- To complete the selected CUSAMS framework (Bolton *et al*., 2004), which enables service organizations to assess the complete value of their 'customer assets' and to understand the influence of marketing instruments on them, we propose to measure different customer perceptions related to price, satisfaction and commitment (see Figure 2 in Chapter 2). Service organizations invest in a diverse array of marketing activities designed to stimulate customer behaviour and thereby influence the financial outcomes of the relationship. Bolton and his colleagues consider the following six categories of marketing instruments: price, service quality programs, direct marketing promotions, relationship marketing instruments (e.g., reward programs), advertising/communications and distribution channels. Each of these six categories of marketing instruments differentially affects relationship duration, service usage, and cross buying of services. They generate revenues (via their effect on individual customer behaviours), and they engender fixed and variable costs.

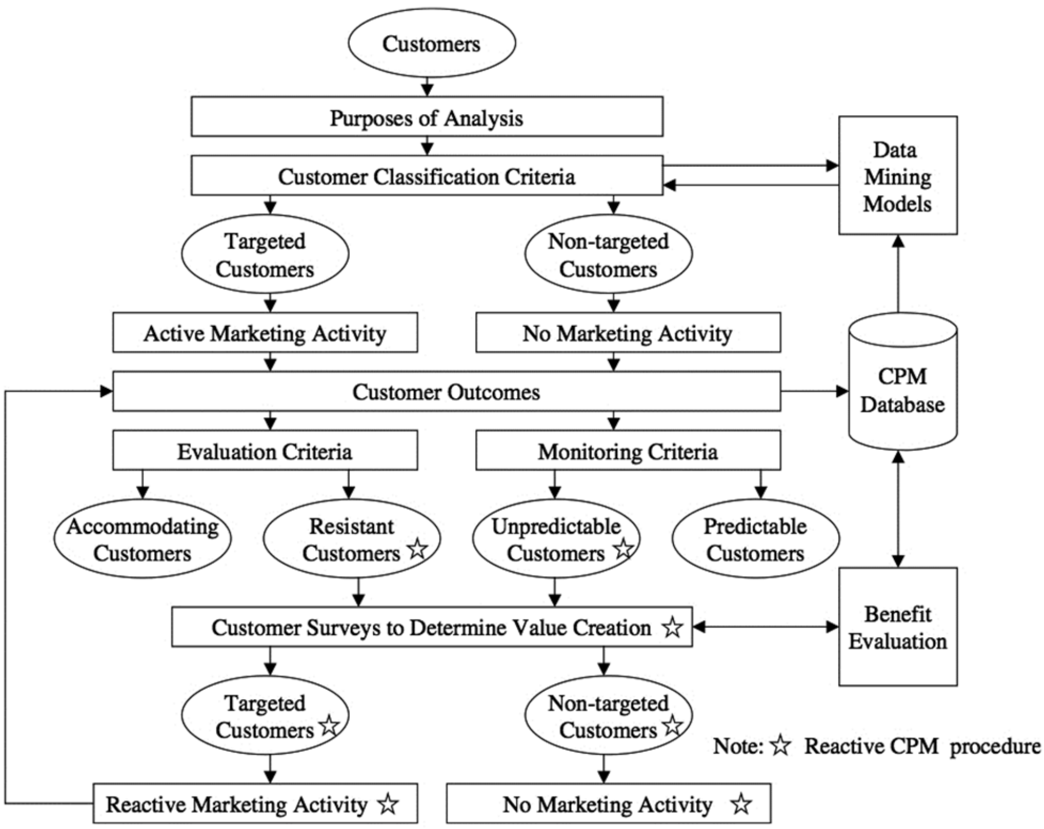
- For future research it is also suggested not to use monthly data to perform the analysis. By using monthly data, one can expect a strong lag (or inertia) effect because customers usually do not decide on bank products on a monthly basis.

Additionally, we include as a future research stream to check the model in the same context but in other countries, also within other economic cycle (not characterised by financial crisis) and finally in other contexts (contractual settings or adapt our model to non-contractual settings as long as we have at our disposal transaction information). The results reported here are based on data provided by a single firm and, therefore, we do not claim to have provided a universally applicable test of the CLV and CE framework

Finally, and taking into account the CLV and CE criticism by Wang and Hong (2006) and Osborne and Ballantyne (2012) (for more details see limitations and conclusions sections, respectively), we remark as a prolific future research stream to develop a wider framework that serves as a guide to get the **"two sided-marketing system and customer centricity"**, combining both concepts: value *from* customer and value *to* customer. We are going to explain below how these authors had taken into account customers as active participants in the value creation process of the firm to develop their models highlighting how these models could be improved with CLV concept.

Wang and Hong (2006) have developed a Customer Profitability Management (CPM) system to overcome CLV management. This CPM system emphasises a continuously interplay between the active and reactive monitoring procedures to identify customers shifts and it is an effective approach to help a firm calibrate its marketing tactics with regard to different types of customers in different situations. It achieves marketing goals by leading customers to migrate along pre-determined and desirable tracks. See Figure 30 below for more details about this system. This is a good example to get the proposed fusion between value *to* customer and value *from* customer, but it is oriented to marketing activities and their impact on customer behaviours. From our point of view this system can be enriched using a more powerful measure to get value from customer, such as our CLV model (the authors used CP).

Figure 30. Active/reactive CPM procedure (*)



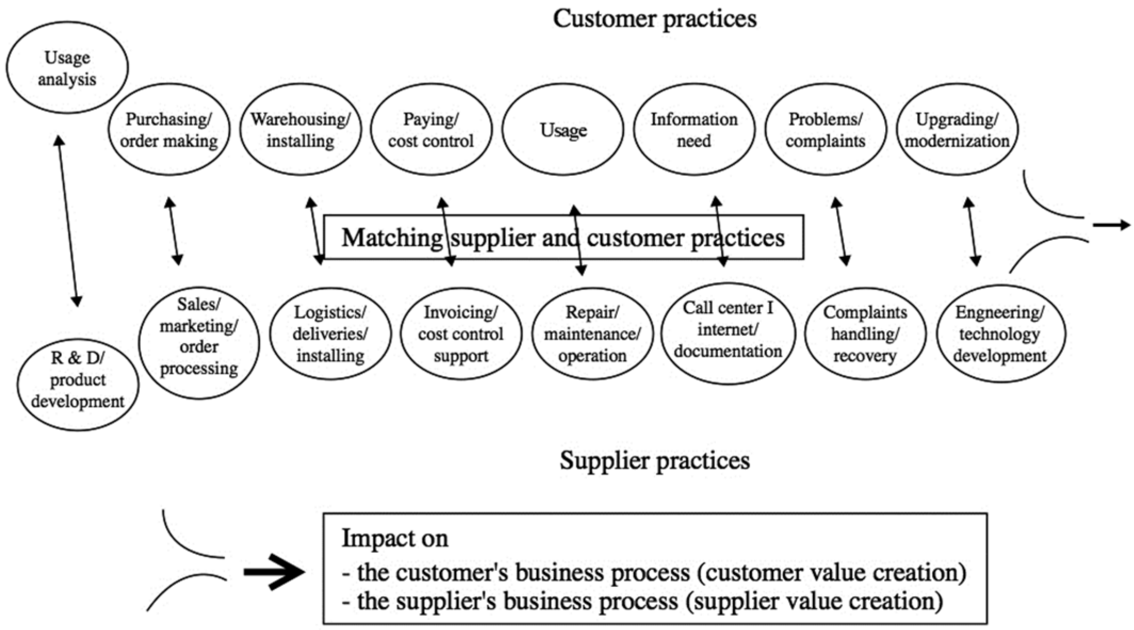
(*) Source: Wang and Hong (2006)

Osborne and Ballantyne (2012) have also posited directions for inquiry to continue looking for the customer centricity, in particular they have proposed: (1) internal marketing, coordinating activities and knowledge resource within and between firms whose realization impacts directly on employee attitudes and the skills needed to meet the needs of customers; (2) service-dominant logic, that encourages reassessment of the role of the customer in the creation of value; (3) collaborating to achieve customer solutions through a step-wise relational process for assisting customers to achieve these mentioned solutions, this moves to customization and resource integration; (4) an initiator-participant marketing perspective, that is, rather than customers and suppliers, there are initiators and participants in any market encounter; and (5) strategy-as-practice, that is a promising new managerial perspective that appears to have the potential to recognise the collaboration required between customers and suppliers to achieve customer solutions.

Additionally, Osborne and Ballantyne (2012) have remarked Grönroos and Helle's (2010) research as a possible way to achieve a **mutual value creation process** within firms[12]. They adopt the service logic in the manufacturing sector for measuring mutually created value in business relationships, which also enables suppliers and customers to share this value between them. Adopting a service logic (see Figure 31) would mean that all activities and processes of a supplier that are relevant to its customer's business are coordinated with the customer's corresponding activities and processes into one integrated stream of actions, with the aim to support the customer's processes and eventually the business outcome. Therefore, the framework includes a conceptual foundation for understanding the process of mutual value creation as well as theoretical basis and measures for calculating mutually created value, joint productivity gains (JPGs) and value sharing.

[12] While Grönroos and Helle (2010) are able to demonstrate how this calculation can be made, applying service logic, the complex nature of assessing mutual value creation remains.

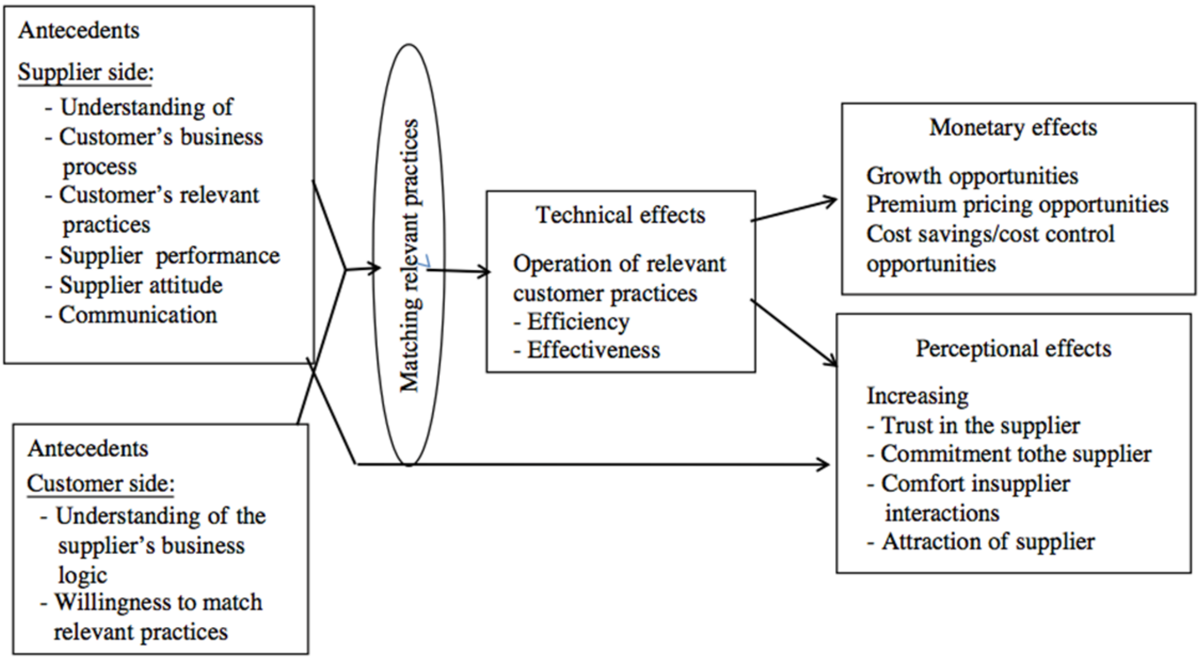
Figure 31. Customer and supplier practices in business-to-business relationships (*)



(*) Source: based on a figure in Grönroos (2008)

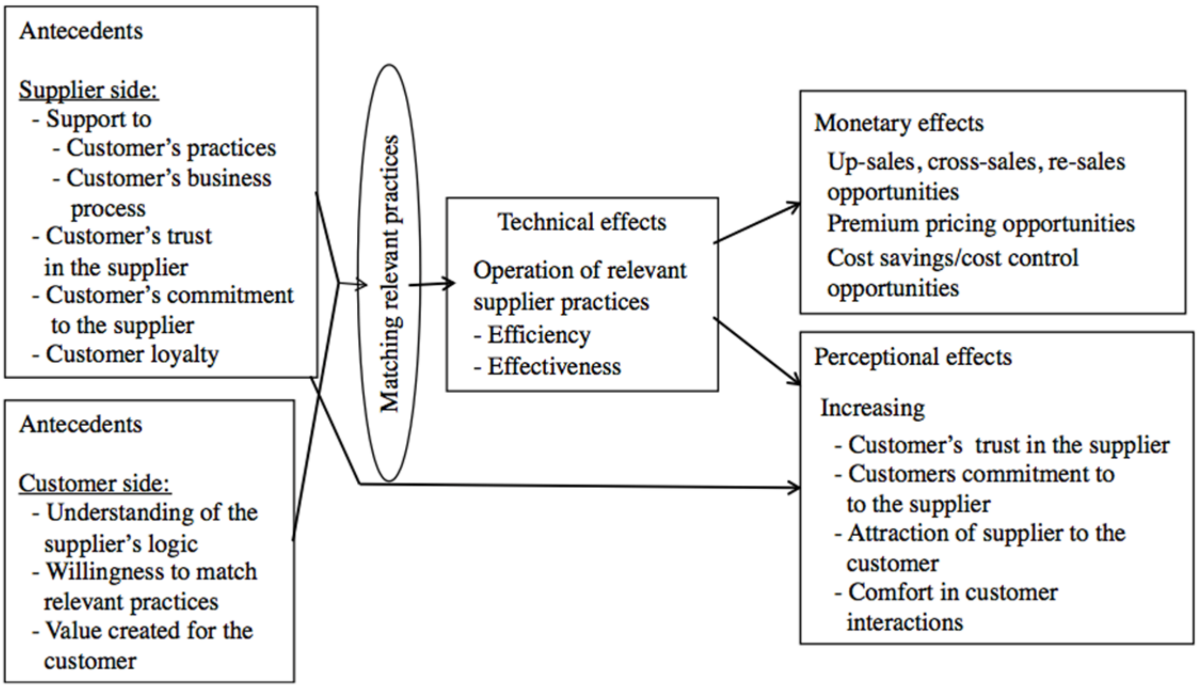
Grönroos and Helle (2010) have considered three value dimensions for the customer and for the supplier, because mutual value creation for the parties requires two models. These three dimensions resemble the functional, economic and psychological dimensions suggested by Gupta and Lehman (2005) in their analysis of customers as investments. In particular, value has a technical dimension, a monetary dimension and also it has a perceptional dimension (i.e., including aspects such as trust, commitment, comfort, attraction) (Holbrook, 1994). The customer's value-creating process (value *for* the customer, see Figure 32) and the supplier's value-creation process (value *for* the supplier, see Figure 33) are connected, because value is generated for both parties from the same business engagement and due to the JPG that can be achieved. The supplier strives to serve its customer by supporting the customer's practices in a way that has a favourable impact on its business, which requires that the customer engages itself in matching its practices with the supplier's corresponding practices in order to get the intended value-creating support. Through a mutual matching of corresponding practices relevant to the customer's business process, resources and competencies on both sides are aligned, which enables the supplier to successfully serve the customer in a value-creating way.

Figure 32. Customer value creation logic (value *for* the customer) (*)



(*) Source: Grönroos and Helle (2010)

Figure 33. Supplier value creation logic (value *for* the supplier) (*)

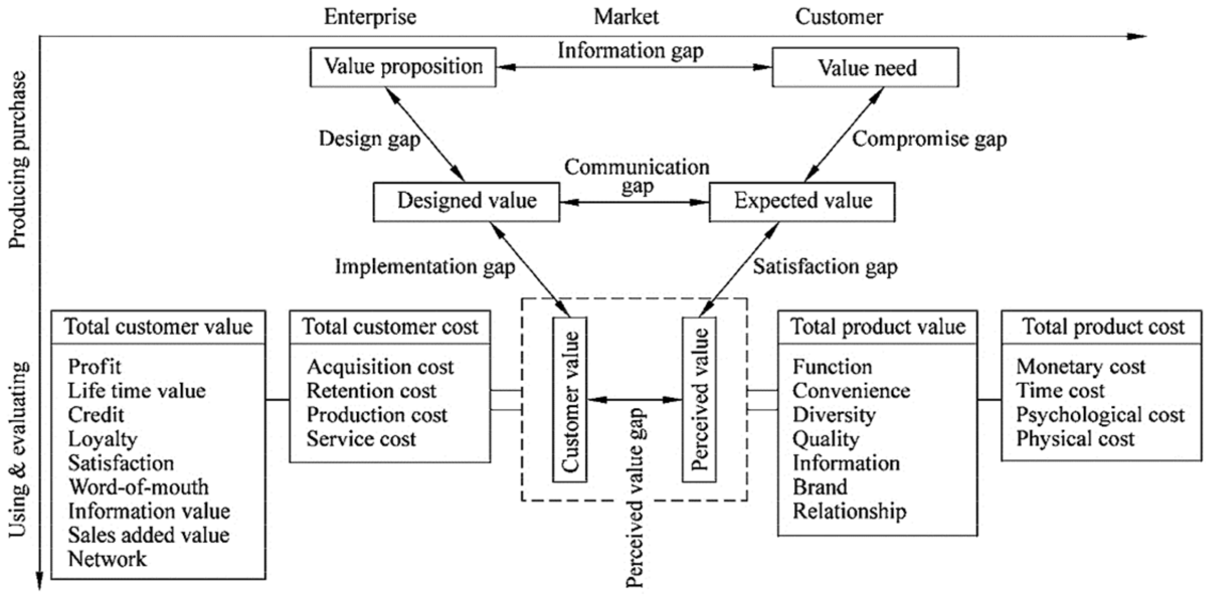


(*) Source: Grönroos and Helle (2010)

In our particular context, that is, banking context, this view highlights the importance of coordinating information within the firm, between customer and firm and across time to manage the entire mutual value system matching their relevant practices. For example, a bank customer who has both a loan product and a saving product might interact with the bank through various channels and different types of interactions (e.g., transactions, information request, complaint), which may change over time. A system on the customer-facing level would capture these interactions and, on the basis of the generated intelligence, would result in coordinated and well-defined actions through different functions. This is translated to get technical effect, monetary effects (e.g., growth, premium prices and cost saving/cost control opportunities) and perceptional effects (e.g., trust, commitment, comfort and attraction to the supplier). On the other hand, the supplier would use this information to get technical effect, monetary effects (e.g., up-sales, cross-sales, re-sales, premium prices opportunities) and perception effects (e.g., customer trust, commitment, attraction, comfort). Therefore, this is another good example to get the proposed fusion between value *to* customer and value *from* customer, but again this system requires of a more powerful measure to get value *from* customer, such as CLV. In this case, the authors develop a set of measures (i.e., mutually created value, joint productivity gains (JPGs) and value sharing) that are not predictive. These measures are a new development from these authors that do not have received high levels of attention. In our opinion, this lack of attention is because JPGs are a less accurate customer value measure than CLV. Therefore, combining the proposed measures in both processes (i.e., value *for* the customer and value *for* the supplier) as antecedents or predictors in our CLV model, the model could be enriched with this wider perspective that we are searching.

As the last example, Lun and Xiaowo (2008) have studied customer value, the key problem of Customer Relationship Management, through a gap model for dual customer values (see Figure 34).

Figure 34. Dual customer value gap model (*)



(*) Source: Lun and Xiaovo (2008)

In this model, as it has been proposed, customer value has two meanings: perceived value (value *to* customer) and customer value (value *from* customer). Related to the first value (value *to* customer), a wide set of perceptions are collected: total product value (functional value, convenience value, diversity value, quality value, information value, brand value and relationship value) and total product cost (monetary cost, time cost, physical cost and psychological cost) (for more details about these measures see Lun and Xiaowo (2008)). On the other hand, about value *from* customer they also collect measures about customer value (customer profit, customer lifetime value, customer credit, loyalty, satisfaction, word-of-mouth, customer information value, sales added-value and customer network value) and about customer cost (customer acquisition cost, production cost, service supply cost). In this article, authors only define and explain these different measures, but without enough details. Therefore, this is an interesting point to start building a model of this kind, knowing that it should be interesting to measure value *to* customer through customer's perceptions and value *from* customer from more objective measures (for example, from a panel data or a company database) and define this second value in a predictive way (using CLV).

These ideas help us to consider a wider perspective about customer value that (if it is possible) we are going to exploit in future research streams combining value for the firm (value *from* customer) and value for the customer (value *to* customer).

REFERENCES

Abe, M. (2009a). “Counting Your Customers One by One: A Hierarchical Bayes Extension to the Pareto/NBD Model”, *Marketing Science*, 28(3), pp. 541-553.

Abe, M. (2009b). “Customer lifetime value and RFM data: Accounting your customers one by one”, *Discussion paper CIRJE-F-616, Graduate School of Economics, University of Tokyo*, available at: http://www.e.u-tokyo.ac.jp/cirje/research/dp/2009/ 2009cf616ab.html.

Acock, A. C. (1997). “Working with missing data”, *Family Science Review*, 10, 76-102.

Ailawadi, K. L.; Lehmann, D. R. and Neslin, S. A. (2001a). “Market response to a major policy change in the marketing mix: learning from Procter & Gamble’s value pricing strategy”, *Journal of Marketing*, 65(1), pp. 44-61.

Ailawadi, K. L.; Neslin, S. A. and Gedenk, K. (2001b). “Pursuing the value-conscious consumer: store brands versus national brand promotions”, *Journal of Marketing*, 65(1), pp. 71-89.

Allenby, G. M.; Leone, R. P. and Jen, L. (1999). “A Dynamic Model of Purchase Timing With Application to Direct Marketing”, *Journal of the American Statistical Association*, 94(June), pp. 365-374.

Allison, P. D. (2001) *Missing data* (Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-136), Sage Publications, Thousand Oaks, California.

Allison, P. D. (2002). *Missing data*, Sage, Thousand Oaks, CA.

Allison, P. D. (2011). *Multiple imputation for missing data: A cautionary tale*, in Vogt, W. P., Sage Quantitative Research Methods, Volume I-IV, Sage Benchmarks in Social Research Methods, Sage publications, Thousand Oaks, CA., pp. 259-268.

Anderson, E. W.; Fornell, C. and Lehmann, D. R. (1994). “Customer Satisfaction, Market Share, and Profitability: Findings from Sweden”, *Journal of Marketing*, 58(3), pp. 53- 66.

Anderson, E. W. and Mittal, V. (2000). “Strengthening the satisfaction-profit chain”, *Journal of Service Research*, 3(2), pp. 107-20.

Apte, C. V.; Hong, S. J.; Natarajan, R.; Pednault, E. P. D.; Tipu, F. A. and Weiss, S. M. (2003). “Data-intensive analytics for predictive modelling”, *IBM Journal of Research & Development*, 47(1), pp. 17-23.

Aravindakshan, A.; Rust, R.; Lemon, K. and Zeithaml, V. A. (2004). “Customer Equity: Making marketing strategy financially accountable”, *Journal of Systems Science and Systems Engineering*, 13(4), pp. 405-422.

Ascarza, E. and Hardie, B. G. S. (2012). “A joint model of usage and churn in contractual settings”, *Working Paper*, available at: http://ssrn.com/abstract=1898552 or http://dx.doi.org/10.2139/ssrn.1898552.

Athanassopoulos, A. D. (2000). “Customer satisfaction cues to support market segmentation and explain switching behaviour”, *Journal of Business Research*, 47(3), pp. 191-207.

Backman, S. J. and Crompton, J. L. (1991). “The Usefulness of Selected Variables for Predicting Activity Loyalty”, *Leisure Science*, 13(3), pp. 205-20.

Baesens, B.; Verstraeten, G.; Van den Poel, D.; Egmont-Petersen, M.; Van Kenhove, P. and Vanthienen, J. (2004). “Bayesian Network classifiers for identifying the slope of customer lifecycle of long-life customers”, *European Journal of Operational Research*, 156(2), pp. 508-523.

Baesens, B.; Viaene, S.; Van den Poel, D.; Vanthienen, J. and Dedene, G. (2002). “Bayesian neural network learning for repeat purchase modelling in direct marketing”, *European Journal of Operational Research*, 138(1), pp. 191-211.

Barney, J. (1991). “Firm Resources and Sustained Competitive Advantage”, *Journal of Management*, 17(1), pp. 99-120

Bauer, H. H. and Hammerschmidt, M. (2005). “Customer-based corporate valuation. Integrating the concepts of customer equity and shareholder value”, *Management Decision*, 43(3), pp. 331-348.

Beane, T.P. and Ennis, D.M. (1987). “Market segmentation: a review”, *European Journal of Marketing*, 21(5), pp. 20-42.

Bell, D.; Deighton, J.; Reinartz, W. J.; Rust, R. T. and Swartz, G. (2002). “Seven barriers to Customer Equity Management”, *Journal of Service Research*, 5(1), pp. 77-85.

Bennett, R. and Rundle, T. (2002). “A comparison of attitudinal loyalty measurement approaches”, *Journal of Brand Management*, 9(3), pp. 193-209.

Benoit, D. F. and Van den Poel, D. (2009). “Benefits of quantile regression for the analysis of customer lifetime value in a contractual setting: An application in financial services”, *Expert Systems with Applications*, 36(7), pp. 10475-10484.

Berger, P. D.; Bolton, R. N.; Bowman, D.; Briggs, E.; Kumar, V.; Parasuraman, A. and Terry, C. (2002). “Marketing actions and the value of customer assets. A framework for customer asset management”, *Journal of Service Research*, 5(1), pp. 39-54.

Berger, P. D.; Eechambadi, N.; George, M.; Lehmann, D. R.; Rizley, R. and Venkatesan, R. (2006). “From Customer Lifetime Value to Shareholder Value. Theory, Empirical evidence, and issues for future research”, *Journal of Service Research*, 9(2), pp. 156-167.

Berger, P. and Magliozzi, T. (1992). “The effect of sample size and proportion of buyers in the sample on the performance of list segmentation equations generated by regression analysis”, *Journal of Direct Marketing*, 6(1), pp. 13-22.

Berger, P. D. and Nasr, N. I. (1998). “Customer lifetime value: marketing models and applications”, *Journal of Interactive Marketing*, 12(1), pp. 17- 30.

Berger, P. D. and Nasr-Bechwati, N. (2001). “The allocation of promotion budget to maximize customer equity”, *Omega*, 29(1), pp. 49-61.

Bhattacharya, C. B. (1998). “When customers are members: Customer retention in paid membership contexts”, *Academy of Marketing Science Journal*, 26(1), pp. 31-44.

Bijak, K. and Thomas, L.C. (2012). “Does segmentation always improve model performance in credit scoring?”, *Expert Systems with Applications*, 39(3), pp. 2433-2442.

Bitran, G. R. and Mondschein, S. (1996). “Mailing decisions in the catalog sales industry”, *Management Science*, 42(9), pp. 1364-1381.

Blattberg, R. C. and Deighton, J. (1996). “Manage marketing by the customer equity test”, *Harvard Business Review*, 74(4), pp.136- 144.

Blattberg, R.; Getz, G. and Thomas, J. S. (2001a). *Customer equity: Building and managing relationships as valuable assets*, Harvard Business School Press, Boston MA.

Blattberg, R.; Getz, G. and Thomas, J. S. (2001b). “Managing customer acquisition”, *Direct Marketing*, 64(6), pp. 41-54.

Blattberg, R. C.; Kim, B. D. and Neslin, S. A. (2008). *Database marketing. Analyzing and managing customers*, Springer, New York.

Blattberg, R. C.; Malthouse, E. and Neslin, S. A. (2009). “Customer Lifetime Value: Empirical generalizations and some conceptual questions”, *Journal of Interactive Marketing*, 23(2), pp. 157-168.

Bloch, M. and Pigneur, Y. (1995). “The extended enterprise: A descriptive framework: Some enabling technologies and case studies in the Lotus Notes Environment”, *Technical Report, University of Lausanne, HEC- INFORGE, TR YP-91*, June.

Bohling, T.; Bowman, D.; LaValle, S.; Mittal, V.; Narayandas, D.; Ramani, G. and Varadarajan, R. (2006). “CRM Implementation: Effectiveness Issues and Insights”, *Journal of Service Research*, 9(2), pp. 184-194.

Bolton, R. N.; Kannan, P. K. and Bramlett, M. D. (2000). “Implications of loyalty program membership and service experiences for customer retention and value”, *Journal of the Academy of Marketing Science*, 28(1), pp. 95-108.

Bolton, R. N. and Lemon, K. N. (1999). “A dynamic model of customers’ usage of services: Usage as an antecedent and consequence of satisfaction”, *Journal of Marketing Research*, 36(2), pp. 171-186.

Bolton, R. N.; Lemon, K. N. and Verhoef, P. C. (2004). “The theoretical underpinnings of customer asset management: a framework and propositions for future research”, *Journal of the Academy of Marketing Science*, 32(3), pp. 271-293.

Borle, S.; Singh, S. and Jain, D.C. (2008). “Customer Lifetime Value measurement”, *Management Science*, 54(1), pp. 100-112.

Bose, R. (2009). “Advanced analytics: opportunities and challenges”, *Industrial Management & Data Systems*, 109(2), pp. 155-172.

Boyce, G. (2000). “Valuing Customers and Loyalty: The Rhetoric of Customer Focus Versus the Reality of Alienation and Exclusion of (Devalued) Customers”, *Critical Perspectives on Accounting*, 11(6), pp. 649-689.

Brijs, T.; Swinnen, G.; Vanhoof, K. and Wets, G. (2004). “Building an association rules framework to improve product assortment decisions”, *Data Mining and Knowledge Discovery*, 8(1), pp. 7-23.

Brooks, S.P. and Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulations, *Journal of Computational and Graphical Statistics*, Vol. 7(4), pp.434-455.

Bruhn, M.; Georgi, D. and Hadwich, K. (2006). “Dimensions and implementation drivers of customer equity management (CEM). Conceptual framework, qualitative evidence and preliminary results of a quantitative study”, *Journal of Relationship Marketing*, 5(1), pp. 21-38.

Buckinx, W. and Van den Poel, D. (2005). “Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting”, *European Journal of Operational Research*, 164(1), pp. 252-268.

Buhl, H. U. and Heinrich, B. (2008). “Valuing Customer Portfolios under Risk-Return-Aspects: A Model-based Approach and its Application in the Financial Services Industry”, *Academy of Marketing Science Review*, 12(5), pp. 1-32.

Bursk, E.C. (1966). “View Your Customers as Investments”, *Harvard Business Review*, 44(5-6), pp. 91-94.

Business World (2007). *Worsening subprime crisis*, November, 27th, pp. 1-5.

Calciu, M. (2008). “Numeric decision support to find optimal balance between customer acquisition and retention spending”, *Journal of Targeting, Measurement and Analysis for Marketing*, 16(3), pp. 214-227.

Calciu, M. (2009). “Deterministic and stochastic Customer Lifetime Value models. Evaluating the impact of ignored heterogeneity in non-contractual contexts”, *Journal of Targeting, Measurement and Analysis for Marketing*, 17(4), pp. 257-271.

Calciu, M. and Salerno, F. (2002). “Customer value modelling: Synthesis and extension proposals”, *Journal of Targeting, Measurement and Analysis for Marketing*, 11(2), pp. 124-147.

Campbell, D. (2006). “The cost structure and customer profitability implications of electronic distribution channels: evidence from online banking”, *Working Paper, Harvard Business School*.

Campbell, D. and Frei, F. (2004). "The persistence of customer profitability: Empirical evidence and implications from a financial services firm", *Journal of Service Research*, 7(2), pp. 107-123.

Campbell, D. and Frei, F. (2010). "Cost Structure, Customer Profitability, and Retention Implications of Self-Service Distribution Channels: Evidence from Customer Behavior in an Online Banking Channel", *Management Science*, 56(1), pp. 4-24.

Chan, C. C. H. (2005). "Online Auction Customer Segmentation Using a Neural Network Model", *International Journal of Applied Science and Engineering*, 3(2), pp. 101-109.

Chan, C. C. H. (2008). "Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer", *Expert Systems with Applications*, 34(4), pp. 2754-2762.

Chandon, P.; Morwitz, V.G. and Reinartz, W.J. (2005). "Do Intentions Really Predict Behavior? Self-Generated Validity Effects in Survey Research", *Journal of Marketing*, 69(2), pp. 1-14.

Chang, W. L. (2011). "iValue: A knowledge-based system for estimating customer prospect value", *Knowledge-Based Systems*, 24(8), pp. 1181–1186.

Chang, H. H. and Tsay, S. F. (2004). "Integrating of SOM and K-mean in data mining clustering: An empirical study of CRM and profitability evaluation", *Journal of Information Management*, 11(4), pp. 161-203.

Ching, W. K.; Ng, M. K. and So, M. M. (2004). "Customer migration, campaign budgeting, revenue estimation: The elasticity of Markov Decision Process on customer lifetime value", *Advanced Modeling and Optimization*, 6(2), pp. 65-80.

Colgate, M. R. and Danaher, P .J. (2000). "Implementing a customer relationship strategy: The asymmetric impact of poor versus excellent execution", *Journal of the Academy of Marketing Science*, 28(3), pp. 375–387.

Colgate, M. R. and Lang, B. (2001). "Switching Barriers in Consumer Markets: An Investigation of the Financial Services Industry", *Journal of Consumer Marketing*, 18(4), pp. 332-47.

Colgate, M.; Stewart, K. and Kinsella, R. (1996). "Customer defection: A study of the student market in Ireland", *International Journal of Bank Marketing*, 14(3), pp. 23-29.

Collins, L. M., Schafer, J. L. and Kam, C. M. (2001). "A comparison of inclusive and restrictive strategies in modern missing data procedures", *Psychological Methods*, 6(4), pp. 330-351.

Collis, D. J. and Montgomery, C. A. (1995). "Competing on resources: Strategy in the 1990s", *Harvard Business School*, 73(4), pp. 118-128.

Colombo, R. and Jiang, W. (1999). "A Stochastic RFM model", *Journal of Interactive Marketing*, 13(3), pp. 2-12.

Courtheoux, R. (1995). *Customer retention: how much to invest. Research and the Customer Lifecycle*, DMA, New York, NY.

Cowles, M. K. (2004). "Review of WinBUGS 1.4", *The American Statistician*, 58(4), pp. 330-336.

Cravens, D. W.; Greenley, G.; Piercy, N. F. and Slater, S. (1997). "Integrating Contemporary Strategic Management Perspectives", *Long Range Planning*, 30(August), pp. 493-506.

Crawley, M. J. (2007). *The R book*, John Willey & Sons, Ltd., The Atrium, Southern Gate, Chichester, West Sussex, England

Crowder, M.; Hand, D.J. and Krzanowski, W. (2007). "On optimal intervention for customer lifetime value", *European Journal of Operational Research*, 183(3), pp. 1550-1559.

Damodaran, A. (2002). *Investment Valuation – Tools and Techniques for Determining the Value of Any Asset*, 2nd Edition, John Wiley & Sons, Inc., New York.

Dawes, J. and Swailes, S. (1999). “Retention sans frontiers: Issues for financial service retailers”, *International Journal of Bank Marketing*, 17(1), pp. 36-43.

Day, G. S. (1969). “A Two-Dimensional Concept of Brand Loyalty”, *Advertising Research*, 9(3), pp. 29-35.

Dick, A. S. and Basu, K. (1994). “Customer loyalty: toward an integrated conceptual framework”, *Journal of the Academy of Marketing Science*, 22(2), pp. 99-113.

Donkers, B.; Verhoef, P. and De Jong, M. (2007). “Modeling CLV: A test of competing models in the insurance industry”, *Quantitative Marketing and Economics*, 5(2), pp. 163-190.

Donkers, B.; Verhoef, P. and De Jong, M. (2003). “Predicting Customer Lifetime Value in Multi-Service Industries”, *ERIM Report Series Reference No. ERS-2003-038-MKT.*, available at: http://ssrn.com/abstract=1290946.

Drew, J.H.; Mani, D.R.; Betz, A.L. and Datta, P. (2001). “Targeting customers with statistical and data-mining techniques”, *Journal of Service Research*, 3(3), pp. 205-219.

Drèze, X. and Bonfrer, A. (2002). “Lifetime value maximization through optimal communication timing”, *Working Paper, Wharton-SMU Research Center.*

Drèze, X. and Bonfrer, A. (2003). “A renewable-resource approach to database valuation”, *Working Paper, Wharton-SMU Research Center.*

Drozdenko, R. G. and Drake, P. D. (2002). *Optimal database marketing: Strategy, development, and data mining*. Sage, London.

Dwyer, F.R. (1989). “Customer Lifetime Valuation to Support Marketing Decision Making”, *Journal of Direct Marketing*, 3(4), pp. 8-15.

Dwyer, F.R. (1997). “Customer Lifetime Valuation to Support Marketing Decision Making”, *Journal of Direct Marketing*, 11(4), pp. 8-15.

Dwyer, F. R.; Schurr, P.H. and Sejo, O. (1987). “Developing buyer-seller relationships”, *Journal of Marketing*, 51(2), pp. 11-27.

Dyche, J. and Tech, J. (2001). *The CRM Handbook: A Business Guide to Customer Relationship Management*, Addison-Wesley, Reading, MA.

Enders, C. K. (2003). “Using the expectation maximization algorithm to estimate coefficient alpha for scales with item-level missing data”, *Psychological Methods*, 8(3), pp. 322-337.

Engel, J.F.; Blackwell, R.D. and Miniard, P.W. (1995). *Consumer Behavior*. The Dryden Press, Orlando, FL.

Etzion, O.; Fisher, A. and Wasserkrug, S. (2005). “E-CLV: A modeling approach for customer lifetime evaluation in e-commerce domains, with an application and case study for online auction”, *Information Systems Frontiers*, 7(4/5), pp. 421-434.

Fader, P. S. and Hardie, B. G. S. (2007). “How to project customer retention”, *Journal of Interactive Marketing*, 21(1), pp. 76-90.

Fader, P. S. and Hardie, B. G. S. (2009). “Probability models for customer-base analysis”, *Journal of Interactive Marketing*, 23(1), pp. 61-69.

Fader, P. S.; Hardie, G. S. and Berger, P. (2004). “Customer-Base Analysis with Discrete-Time Transaction Data”, available at: http://brucehardie.com/ papers/020/.

Fader, P. S.; Hardie, B. G. S. and Jerath, K. (2007). “Estimating CLV using aggregated data: The Tuscan lifestyles case revisited”, *Journal of Interactive Marketing*, 21(3), pp. 55-71.

Fader, P. S.; Hardie, B. G. S. and Lee, K. L. (2005a). "Counting your customers the easy way: An alternative to the Pareto/NBD model", *Marketing Science*, 24(2), pp. 275-284.

Fader, P. S.; Hardie, B. G. S. and Lee, K. L. (2005b). "RFM and CLV: Using iso-value curves for customer base analysis", *Journal of Marketing Research*, 42(4), pp. 415-430.

Fader, P. S.; Hardie, B. G. S. and Lee, K. L. (2006). "More than meets the eye", *Marketing Research*, 18(Summer), pp. 9-14.

Farley, J. U. (1964). "Why does brand loyalty vary over products", *Journal of Marketing Research*, 1(4), pp. 9-14.

Felvey, J. (1982). "Cross-selling by computer", *Bank Marketing*, (July), pp. 25-27.

Ganesh, J.; Arnold, M. J.; Reynolds, K. E. (2000). "Understanding the customer base of service providers: An examination of the differences between switchers and stayers", *Journal of Marketing*, 64(3), pp. 65-87.

Garland, R. (2002). "Non-financial drivers of customer profitability in personal retail banking", *Journal of Targeting, Measurement and Analysis for Marketing*, 10(3), pp. 233-248.

Garland, R. (2004). "Share of wallet's role in customer profitability", *Journal of Financial Services Marketing*, 8(3), pp. 259-268.

Gelman, A.; Carlin, J. B.; Stern, H.S . and Rubin, D. B. (1995). *Bayesian Data Analysis*, Chapman & Hall, Boca Raton, FL.

Gelman, A.Y. and Rubin D. (1992). "Inference from iterative simulation using multiple sequences", *Statistical Science*, Vol. 7(4), pp. 457-511.

Geman, S. and Geman, D. (1984). "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAM 1-6, pp. 721-741.

Glady, N.; Baesens, B. and Croux, C. (2009a). "Modeling churn using customer lifetime value", *European Journal of Operational Research*, 197(1), pp. 402-411.

Glady, N.; Baesens, B. and Croux, C. (2009b). "A modified Pareto/NBD approach for predicting customer lifetime value", *Expert Systems with Applications*, 36(2 part 1), pp. 2062-2071

Gleaves, R.; Burton, J.; Kitshoff, J.; Bates, K. and Whittington, M. (2008). "Accounting is from Mars, marketing is from Venus: establishing common ground for the concept of customer profitability", *Journal of Marketing Management*, 24(7-8), pp. 825-845.

Graham, J. W. (2009). "Missing data analysis: making it work in the real world", *Annual Review of Psychology*, 60(1), pp. 549-576.

Grant, R.M. (1991). "The Resource-Based Theory of Competitive Advantage: Implications for Strategy Formulation", *California Management Review*, 33(3), pp. 114-135

Grewal, D.; Levy, M. and Lehmann, D.R. (2004). "Retail branding & customer loyalty: An overview. Editorial", *Journal of Retailing*, 80(4), pp. ix–xii.

Grönroos, C. (2008). "Service logic revisited: who creates value? And who co-creates?", *European Business Review*, 20(4), pp. 298-314.

Grönroos, C. and Helle, P. (2010). "Adopting a service logic in manufacturing: Conceptual foundation and metrics for mutual value creation", *Journal of Service Management*, 21(5), pp. 564-590.

Grover, R. and Vriens, M. (2006). *The Handbook of Marketing Research: Uses, Misuses, and Future Advances*, Sage Publications, Thousand Oaks, California.

Gupta, S. (2009). "Customer-Based Valuation", *Journal of Interactive Marketing*, 23(2), pp. 169-178.

Gupta, S.; Hanssens, D.; Hardie, B.; Kahn, W.; Kumar, V.; Lin, N.; Ravishanker, N. and Sriram, S. (2006). "Modeling Customer Lifetime Value", *Journal of Service Research*, 9(2), pp. 139-155.

Gupta, S. and Lehmann, D.R. (2003). "Customers as assets", *Journal of Interactive Marketing*, 17(1), pp. 9-24.

Gupta, S. and Lehman, D. R. (2005). *Managing Customers as Investments*, Wharton School, Upper Saddle River, NJ.

Gupta, S. and Lehmann, D. R. (2008). *Models of Customer Value*, in Wierenga, B. (2008). Handbook of Marketing Decision Models. International Series in Operations Research & Management Science, 121(part IV), pp. 255-290, Springer, Rotterdam, US.

Gupta, S.; Lehmann, D. R. and Stuart, J. A. (2004). "Valuing Customers", *Journal of Marketing Research*, 41(1), pp. 7-18.

Gupta, S. and Zeithaml, V. (2006). "Customer metrics and their impact on financial performance", *Marketing Science*, 25(6), pp. 718-739.

Gurau, C. and Ranchhod, A. (2002). "How to calculate the value of a customer", *Journal of Targeting, Measurement and Analysis for Marketing*, 10(3), pp. 20-220.

Haenlein, M.; Kaplan, A. M. and Beeser, A. J. (2007). "A model to determine customer lifetime value in retail banking context", *European Management Journal*, 25(3), pp. 221-234.

Haentjens, M. (2011). *Predictive uncertainty in customer lifetime value (CLV) predictions. How can uncertainty estimates be incorporated in the customer portfolio approach?*, Master Dissertation Universiteit Gent (Belgium).

Hair J. F.; Black, W. C.; Babin, B. J. and Anderson, R. E. (2009). *Multivariate Data Analysis*, 7th Edition, Prentice-Hall, Upper Saddle River, NJ.

Han, S. L. and Baek, S. B. (2004). "Antecedents and consequences of service quality in online banking: An application of the SERVQUAL instrument", *Advances in Consumer Research*, 31, pp. 208-214.

Hawkins, D.M. and Kass, G.V. (1982). *Automatic interaction detection*, in Hawkins, D.M., *Topics in applied multivariate analysis*, Cambridge University Press: Cambridge, pp. 269-302.

Hedderley, D. and Wakeling, I. (1995). "A comparison of imputation techniques for internal preference mapping using Monte Carlo simulation", *Food Quality and Preference*, 6(4), pp. 281-297.

Hepworth, M. and Mateus, P. (1994). "Connecting Customer Loyalty to the Bottom Line", *Canadian Business Review*, 21(4), pp. 40-44.

Hitt, L. M. and Frei, F. X. (2002). "Do better customers utilize electronic distribution channels? The case of PC Banking", *Management Science*, 48(6), pp. 732-748.

Hoffmann-Jørgensen, J. (1994). *Probability with a view towards statistics*, Chapman & Hall, New York.

Hogan, J.E.; Lemon, K. N. and Libai, B. (2003). "What is the value of a lost customer?", *Journal of Service Research*, 5(3), pp. 196-208.

Hogan, J. E.; Lemon, K. N. and Rust, R. T. (2002). "Customer Equity Management: Charting new directions for the future of Marketing", *Journal of Service Research*, 5(1), pp. 4-12

Holbrook, M. B. (1994). *The nature of customer value – an axiology of services in the consumption experience*, in Rust, R.T. and Oliver, R.L. (Eds), Service Quality: New Directions in Theory and Practice, Sage, Thousand Oaks, CA.

Hsieh, K. L. (2004). "Applying an expert system into constructing customer's value expansion and prediction model based on AI techniques in leisure industry", *Expert Systems with Applications*, 36(2), pp. 2864-2872.

Hu, W. and Jing, Z. (2008). “Study of Customer Segmentation for Auto Services Companies Based on RFM Model”, *Proceedings for ICIM 2008: International Conference on Innovation and Management*, UNU-Maastricht Economic and Social Research Institute on Innovation and Technology, Maastricht, (The Netherlands).

Hughes, A. M. (1996a). “Boosting response with RFM”, *American Demographics (supplement Marketing Tools)*, May, pp. 4-10.

Hughes, A. M. (1996b). *The Complete Database Marketer*, Irwin, Chicago.

Hui-min, M.A.; Jian-ling, C. and Kai, Z. (2006). “Customer value prediction model based on Markov Chain in B2C”, *Proceedings for International Conference on Management Science and Engineering*.

Hunt, S. D. (2000). *A General Theory of Competition: Resources, Competences, Productivity, Economic Growth*, Sage Publications, Thousand Oaks, CA.

Hwang, H.; Jung, T. and Suh, E. (2004). “An LTV model and customer segmentation based on customer value: A case study on the wireless telecommunication industry”, *Expert Systems with Applications*, 26(2), pp. 181-188.

IBM (2010). “Gestión de clientes 2012. Un Nuevo modelo para las entidades financieras”, available at: http://www-304.ibm.com/easyaccess/sectorfinanciero/forms/!!/xmlid=191789.

IBM (2011). *IBM SPSS Missing Values 20*, IBM Corporation, USA.

Jackson, B. B. (1985). *Winning and keeping industrial customers: The dynamics of customer relationships*, Lexington Books, Lexington, MA.

Jackson, D. (1989a). “Determining a customer’s lifetime value”, *Direct Marketing*, 51(11), pp. 60-62.

Jackson, D. (1989b). “Determining a customer’s lifetime value (Part 2)”, *Direct Marketing*, 52(1), pp. 24-32.

Jackson, D. (1989c). “Insurance marketing: Determining a customer’s lifetime value (Part 3)”, *Direct Marketing*, 52(4), pp. 28-30.

Jackson, D. R. (1994). “Strategic application of customer lifetime value in the direct marketing environment”, *Journal of Targeting Measurement and Analysis for Marketing*, 3(1), pp. 9- 17.

Jacobs, F. A.; Johnston, W. and Kotchetova, N. (2001). “Customer profitability: Prospective vs. retrospective approaches in a business-to-business setting”, *Industrial Marketing Management*, 30(4), pp. 353-363.

Jacoby, J. and Chestnut, R. W. (1978). *Brand Loyalty: Measurement and Management,* John Wiley, New York.

Jacoby, J. and Kyner, D. B. (1973). “Brand loyalty vs. repeat buying behaviour”, *Journal of Marketing Research*, 10(1), pp. 1-9.

Jain, D. and Singh, S. (2002). “Customer Lifetime Value Research in Marketing: A Review and Future Directions”, *Journal of Interactive Marketing*, 16(2), pp. 34-46.

Jarvis, C. B.; MacKenzie, S. B. and Podsakoff, P. M. (2003). “A critical review of construct indicators and measurement model misspecification in marketing and consumer research”, *Journal of Consumer Research*, 30(2), pp. 199-218.

Johnson, M. and Selnes, F. (2004). “Towards a dynamic theory of exchange relationships: Customer portfolio management”, *Journal of Marketing*, 68(2), pp. 1-17.

Kahan, R. (1998). “Using database marketing techniques to enhance your one-to-one marketing initiatives”, *Journal of Consumer Marketing*, 15(5), pp. 491-493.

Kalbfleisch, J. and Prentice, R. (1980). *Statistical Analysis of Failure Time Data*, John Wiley and Sons, New York.

Kamakura, W.; Ramaswami, S. and Srivasta, R. (1991). "Applying latent trait analysis in the evaluation of prospects for cross-selling of financial services", *International Journal of Research in Marketing*, 8(4), pp. 329-349.

Kamakura, W. A.; Wedel, M; Rosa F. D. and Mazzon, J. A. (2003). "Cross-selling through database marketing: A mixed data factor analyzer for data augmentation and prediction", *International Journal of Research in Marketing*, 20(1), pp. 45-65.

Kaplan, R. S. and Narayanan, V. G. (2001). "Measuring and managing customer profitability", *Cost Management*, 15(5), pp. 5-15.

Keiningham, T.L.; Aksoy, L. and Bejou, D. (2006). "Approaches to the Measurement and Management of Customer Value", *Journal of Relationship Marketing*, 5(2-3), pp. 37-54.

Kéry, M. (2010). *Introduction to WinBUGS for Ecologist. A Bayesian Approach to Regression, Anova, Mixed Models, and Related Analyses*, Academic Press, Burlington.

Khajvand, M.; Zolfaghar, K.; Ashoori, S. and Alizadeh, S. (2011). "Estimating customer lifetime value based on RFM analysis of customer purchase behaviour: case study", *Procedia Computer Science*, 3, pp. 57-63.

Kim, S. Y.; Jung, T. S.; Suh, E. H. and Hwang, H. S. (2006). "Customer segmentation and strategy development based on customer lifetime value: A case study", *Expert Systems with Applications*, 31(1), pp. 101-107.

Knott, A.; Hayes, A. and Neslin, S. (2002). "Next-product-to-buy models for cross-selling applications", *Journal of Interactive Marketing*, 16(3), pp. 59-75.

Kotler, P. (1974). "Marketing during periods of shortage", *Journal of Marketing*, 38(3), pp. 20-29.

Krasnikov, A.; Jayachandran, S. and Kumar, V. (2009). "The impact of customer relationship management implementation on cost and profit efficiencies: evidence from the U.S. Commercial Banking Industry", *Journal of Marketing*, 73(6), pp. 61-76.

Kumar, V., Petersen, J. A. and Leone, R. P. (2010). "Driving Profitability by Encouraging Customer Referrals: Who,When, and How", *Journal of Marketing*, 74(5), pp. 1-17.

Kumar, V. (2008a). *Managing customers for profit. Strategies to increase loyalty and build profits*. Wharton School Publishing, New Jersey.

Kumar, V. (2008b). "Customer Lifetime Value. The Path to Profitability", *Foundations and Trends in Marketing*, 2(1), pp. 1-96.

Kumar, V. (2006). *Customer Lifetime Value*, in The Handbook of Marketing Research: Uses, Misuses, and Future Advances, Eds. Grover, R. and Vriens, M., pp. 602-627, Sage Publications, Thousand Oaks, CA.

Kumar, V. and George, M. (2007). "Measuring and Maximizing Customer Equity: A Critical Analysis", *Journal of the Academy of Marketing Science*, 35(2), pp. 157-171.

Kumar, V.; George, M. and Pancras, J. (2008a). "Cross-buying in retailing: Drivers and consequences", *Journal of Retailing*, 84(1), pp. 15-27.

Kumar, V.; Petersen, J. A. and Leone, R. P. (2007). "How valuable is word of mouth?", *Harvard Business Review*, October, pp. 1-9.

Kumar, V.; Pozza, I. D.; Petersen, J.A. and Shah, D. (2009). "Reversing The Logic: The Path To Profitability", *Journal of Interactive Marketing*, 23(2), pp. 147-156.

Kumar, V. and Rajan, B. (2009). "Nurturing the right customers", *Strategic Finance*, September, pp. 27-33.

Kumar, V.; Ramani, G. and Bohling, T. (2004). "Customer lifetime value approaches and best practice applications", *Journal of Interactive Marketing*, 18(3), pp. 60-72.

Kumar, V. and Shah, D. (2004). “Building and sustaining profitable customer loyalty for the 21st century”, *Journal of Retailing*, 80(4), pp. 317-330.

Kumar, V. and Shah, D. (2009). “Expanding the role of marketing: from customer equity to marketing capitalization”, *Journal of Marketing*, 73(6), pp. 119-136.

Kumar, V.; Venkatesan, R.; Bohling, T. And Beckmann, D. (2008b). “The power of CLV: Managing Customer Lifetime Value at IBM”, *Marketing Science*, 27(4), pp. 585-599.

Kumar, V.; Shah, D. and Venkatesan, R. (2006a). “Managing retailer profitability – one customer at a time!”, *Journal of Retailing*, 82(4), pp. 277-294.

Kumar, V.; Venkatesan, R. and Reinartz, W. (2006b). “Knowing What to Sell, When, and to Whom”, *Harvard Business Review*, March, pp. 131-137.

Landerman, L. R.; Land, K. C. and Pieper, C. F. (1997). “An Empirical Evaluation of the Predictive Mean Matching Method for Imputing Missing Values”, *Sociological Methods & Research*, 26(1), pp. 3-33.

LaPlaca, P. J. (2004). “Letter from the editor: Special issue on customer relationship management”, *Industrial Marketing Management*, 33(6), pp. 463−464.

Lee, J. H. and Park, S. C. (2005). “Intelligent profitable customers segmentation system based on business intelligence tools”, *Expert Systems with Applications*, 29(1), pp. 145-152.

Lemon, K. N. and Mark, T. (2006). “Customer Lifetime Value as the basis of customer segmentation”, *Journal of Relationship Marketing*, 5(2), pp. 55-69.

Lemon, K. N.; Rust, R. T. and Zeithaml, V. A. (2001). “What drives customer equity?”, *Marketing Management*, 10(1), pp. 20-25.

Leverin, A. and Liljander,V. (2006). “Does relationship marketing improve customer relationship satisfaction and loyalty?”, *International Journal of Bank Marketing*, 24(4), pp. 232-251.

Levitt, T. (1983). “After the Sale is Over...”, *Harvard Business Review*, 61(September/October), pp. 87-93.

Lewis, M. (2005). “A dynamic programming approach to customer relationship pricing”, *Management Science*, 51(6), pp. 980-994.

Lewis, M. (2006). “Customer acquisition promotions and customer assets value”, *Journal of Marketing Research*, 43(2), pp. 195-203.

Li, D. C.; Dai, W. L. and Tsen, W. T. (2011). “A two-stage clustering method to analyze customer characteristics to build discriminative customer management: A case of textile manufacturing business”, *Expert Systems with Applications*, 38(6), pp. 7186-7191.

Liang, Y.H. (2010). “Integration of data mining technologies to analyze customer value for the automotive maintenance industry”, *Expert Systems with Applications*, 37(12), pp. 7489-7496.

Libai, B.; Muller, E. and Peres, R. (2009). “The diffusion of services”, *Journal of Marketing Research*, 46(2), pp. 163-175.

Libai, B.; Narayandas, D. and Humby C. (2002). “Toward and individual customer profitability model: A segment-based approach”, *Journal of Service Research*, 5(1), pp. 69-76.

Ling, R., and Yen, D. C. (2001). “Customer relationship management: An analysis framework and implementation strategies”, *Journal of Computer Information Systems*, 41(3), pp. 82-97.

Lingle, S. (1995). “How much is a customer worth?”, *Bank Marketing*, 27(8), pp. 13-16.

Little, R. J. A. (1988). “A test of missing completely at random for multivariate data with missing values”, *Journal of the American Statistical Association*, 83 (404), pp. 1198-1202.

Little, R. J. and Rubin, D. B. (1989). “The Analysis of Social Science Data with Missing Values”, *Sociological Methods & Research*, 18(2-3), pp. 292-326.

Liu, W. J.; Duan, S. Y.; Yang, X. and Wang, X. F. (2011). “Determination of customer value measurement model RFM index weights”, *African Journal of Business Management*, 5(14), pp. 5567-5572

Liu, D. R. and Shih, Y. Y. (2005a). “Integrating AHP and datamining for product recommendation based on customer lifetime value”, *Information & Management*, 42(3), pp. 387-400.

Liu, D. R. and Shih, Y. Y. (2005b). “Hybrid approaches to product recommendation based on customer lifetime value”, *Information & Management*, 77(2), pp. 181-191.

Lun, H. and Xiaowo, T. (2008). “Gap model for dual customer value”, *Tsinghua Science and Technology*, 13(3), pp. 395-399.

Malthouse, E. C. and Blattberg, R. C. (2005). “Can we predict customer lifetime value?”, *Journal of Interactive Marketing*, 19(1), pp. 2-16.

Malthouse, E. C. and Blattberg, R. C. (2002). “Are Customer Relationship Management Strategies Possible?”, *Working Paper, Northwestern University, Evanston, IL.*

Malthouse, E. and Mulhern, F. (2008). “Understanding and Using Customer Loyalty and Customer Value”, *Journal of Relationship Marketing*, 6(3/4), pp. 59-86.

Marcus, C. (1998). “A practical yet meaningful approach to customer segmentation”, *Journal of Consumer Marketing*, 15(5), pp. 494-504.

Mason, C. H. (2003). “Tuscan Lifestyles: Assessing Customer Lifetime Value”, *Journal of Interactive Marketing*, 17(4), pp. 54-60.

Massey, W. F.; Montgomery, D. B. and Morrison, D. G. (1970). *Stochastic Models of Buyer Behavior*, Cambridge, MIT Press, Massachusetts.

McCarty, J.A. and Hastak, M. (2007). “Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression”, *Journal of Business Research*, 60(6), pp. 656-662.

Metropolis, N.; Rosenbluth, A.; Rosenbluth, M.; Teller, A. and Teller, E. (1953). “Equations of State Calculations by Fast Computing Machines”, *Journal of Chemical Physics*, 21, pp. 1087-1091.

Miglautsch, J. R. (2000). “Thoughts on RFM scoring”, *Journal of Database Marketing*, 8(1), pp. 67-72.

Mittal, V. and Kamakura, W. A. (2001). “Satisfaction, repurchase intent, and repurchase behaviour: Investigating the moderating effect of customer characteristics”, *Journal of Marketing Research*, 38(1), pp. 131-142.

Mittal, V. and Sarkees, M. (2006). “Customer divestment”, *Journal of Relationship Marketing*, 5(2), pp. 71-85.

Mitussis, D.; O’Malley, L. and Patterson, M. (2006). “Mapping the Re-engagement of CRM with Relationship Marketing”, *European Journal of Marketing*, 40(5/6), pp.572- 589.

Morgan, R. M. and Hunt, S. D. (1994). “The commitment–trust theory of relationship marketing”, *Journal of Marketing*, 58(3), pp. 20-38.

Morrison, D. G. and Schmittlein, D. C. (1980). “Jobs, Strikes, and Wars: Probability Models for Duration”, *Organizational Behavior and Human Performance*, 25(2), pp. 224-251.

Mozer, M. C.; Wolniewicz, R.; Grimes, D. B.; Johnson, E. and Kaushansky, H. (2000). "Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry", *IEEE Transactions on Neural Networks*, 11(3), pp. 690-696.

MSI (2004). *Research priorities 2004–2006*, Marketing Science Institute, Boston, MA.

Mulhern, F.J. (1999). "Customer profitability analysis: Measurement, concentration, and research directions", *Journal of Interactive Marketing*, 13(1), pp. 25-40.

Musil, C. M.; Warner, C. B.; Yobas, P. K. and Jones, S. L. (2002). "A comparison of imputation techniques for handling missing data", *Western Journal of Nursing Research*, 2002, 24(7), pp. 815-829.

Mzoughia, M.B. and Limam, M. (2012). "Modelling customer lifetime value in retail banking context", *Proceedings for the MSDM (Meeting on Statistics and Data Mining)*, pp. 33-38.

Neal, R. (1997), "Markov Chain Monte Carlo Methods Based on 'Slicing' the Density Function", *Technical Report 9722, Department of Statistics, University of Toronto, Canada*, available at http://www.cs.toronto.edu/~radford/ftp/slice.pdf.

Neslin, S.; Gupta, S.; Kamakura, W.; Lu, J. and Mason, C. (2006). "Defection detection: Measuring and understanding the predictive accuracy of customer churn models", *Journal of Marketing Research*, 46(May), pp. 204-211.

Ngai, E. W. T.; Xiu, L. and Chau, D. C. K. (2009). "Application of data mining techniques in customer relationship management: A literature review and classification", *Expert Systems with Applications*, 36(2), pp. 2592–2602

Noone, B. and Griffin, P. (1997). "Enhancing yield management with customer profitability analysis", *International Journal of Contemporary Hospitality Management*, 9(2), pp. 75-79

Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS*, John Wiley & Sons, Inc., USA.

Oliver, R. L. (1999). "Whence Consumer Loyalty?", *Journal of Marketing*, 63(Special Issue), pp. 33-44.

Osborne, P. and Ballantyne, D. (2012). "The paradigmatic pitfalls of customer-centric marketing", *Marketing Theory*, 12(2) 155-172.

Paas, L. J.; Bijmolt, T. H. A. and Vermunt, J. K. (2007). "Acquisition patterns of financial products: A longitudinal investigation", *Journal of Economic Psychology*, 28(2), pp. 229-241.

Paas, L. J. and Molenaar, I. W. (2005). "Analysis of acquisition patterns: A theoretical and empirical evaluation of alternative methods", *International Journal of Research in Marketing*, 22(1), pp. 87-100.

Paauwe, P.; Van der Putten, P. and Van Wezel, M. (2007). "DTMC: An actionable e-customer lifetime value model based on Markov Chains and Decision Trees", *Proceedings for the 9th International Conference on Electronic Commerce*.

Parvatiyar, A. and Sheth, J. N. (2001). "Customer relationship management: Emerging practice, process, and discipline", *Journal of Economic and Social Research*, 3(2), pp. 1-34.

Paulin, M.; Perrien, J.; Ferguson, R.J.; Salazar, A. M. A. y Seruya, L. M. (1998). "Relational norms and client retention: External effectiveness of commercial banking in Canada and Mexico", *International Journal of Bank Marketing*, 16(1), pp. 24–31.

Payne, A. and Frow, P. (2005). "A strategic framework for customer relationship management", *Journal of Marketing*, 69(4), pp. 167-176.

Pearson, S. (1996). *Building brands directly: creating business value from customer relationships*, MacMillan Business, London.

Persson, A. and Ryals, L. (2010). “Customer assets and customer equity: Management and measurement issues”, *Marketing Theory*, 10(4), pp. 417-436.

Petrison, L. A; Blattberg, R. C. and Wang, P. (1993). “Database marketing: Past, present and future”, *Journal of Direct Marketing*, 7(3), pp. 27-43.

Peugh, J. L. and Enders, C. K. (2004). “Missing data in educational research: A review of reporting practices and suggestions for improvement”, *Review of Educational Research*, 74(4), pp. 525-556.

Pfeifer, P. and Carraway, R. (2000). “Modeling customer relationships as Markov chains”, *Journal of Interactive Marketing*, 14(2), pp. 43-55.

Pfeifer, P. E.; Haskins, M. E. and Conroy, R. M. (2005). “Customer Lifetime Value, Customer Profitability, and the treatment of acquisition spending”, *Journal of Managerial Issues*, 17(1), pp. 11-25.

Prinzie, A. and Van den Poel, D. (2006). “Incorporating sequential information into traditional classification models by using an element/position-sensitive SAM”, *Decision Support Systems*, 42(2), pp. 508-526.

Prinzie, A. and Van den Poel, D. (2007). “Predicting home-appliance acquisition sequences: Markov/Markov for Discrimination and survival analysis for modelling sequential information in NPTB models”, *Decision Support Systems*, 44(1), pp. 28-45.

Prinzie, A. and Van den Poel, D. (2008). “Modeling complex longitudinal consumer behavior with Dynamic Bayesian networks: an Acquisition Pattern Analysis application”, *Journal of Intelligent Information Systems*, 36(6), pp. 283-304.

Pritchard, M. P. and Howard, D. R. (1997). “The Loyal Travel: Examining a Typology of Service Patronage”, *Travel Research*, 35(4), pp. 2-10.

Ramaswamy, V.; Raghunathan, T. E.; Cohen, S. H. and Ozcan, K. (2001). “A multiple imputation approach for the analysis of missing data in marketing research”, *Unpublished working paper, University of Michigan, Department of Business Administration*.

Rappaport, A. (1986). *Creating Shareholder Value: The New Standard for Business Performance*, The Free Press, USA.

Rasmusson, E. (1999). “Complaints can build relationships”, *Sales and Marketing Management*, 151(9), pp. 89-90.

Reichheld, F. F. (1996). *The Loyalty Effect: The Hidden Force Behind Growth, Profits and Lasting Value*, Harvard Business School Press, Boston, MA.

Reichheld, F. F. and Sasser, W.E. (1990). “Zero defections: Quality comes to services”, *Harvard Business Review*, 68(5), pp. 105-111.

Reichheld, F. F. and Teal, T. (1996). *The loyalty effect: The hidden force behind growth, profits, and lasting value*, Harvard Business School Press, Boston, Massachusetts.

Reinartz, W. J. and Kumar, V. (2000). “On the profitability of long-life customers in a non-contractual setting: An empirical investigation and implications for marketing”, *Journal of Marketing*, 64(4), pp.17-35.

Reinartz, W. and Kumar, V. (2002). “The mismanagement of customer loyalty”, *Harvard Business Review*, 80(7), pp. 86-94.

Reinartz, W. and Kumar, V. (2003). “The impact of customer relationship characteristics on profitable lifetime duration”, *Journal of Marketing*, 67(1), pp. 77-99.

Reinartz, W.; Krafft, M. and Hoyer, W.D. (2004). “The Customer Relationship Management Process: Its Measurement and Impact on Performance”, *Journal of Marketing Research*, 41(3), pp. 293-305.

Reinartz, W.; Thomas, J. S. and Bascoul, G. (2008). "Investigating cross-buying and customer loyalty", *Journal of Interactive Marketing*, 22(1), pp. 5-20.

Reinartz, J. S; Thomas, V.; Kumar. (2005). "Balancing Acquisition and Retention Resources to Maximize Customer Profitability", *Journal of Marketing*, 69(1), pp. 63-79.

Reinartz, W. and Venkatesan, R. (2008). *Decision Models for Customer Relationship Management (CRM),* in Wierenga, B. (2008). Handbook of Marketing Decision Models, International Series in Operations Research & Management Science, Volume 121, Part IV, pp. 291-326, Springer, Rotterdam, US.

Richards, K. A. and Jones, E. (2008). "Customer relationship management: finding value drivers", *Industrial Marketing Management*, 37(2), pp.120–130.

Rigby, D. K.; Reichheld, F. F. and Schefter, P. (2002). "Avoid the four perils of CRM", *Harvard Business Review*, 80(2), pp. 101-109.

Ritter, D. S. (1993). *Relationship Banking*, Probus Publishing Company, Cambridge.

Robert, J. (2007). *The Bayesian choice: From decision-theoretic foundations to computational implementation*, Springer, NY.

Roberts, M. L. and Berger, P. D. (1989). *Direct marketing management*, Prentice-Hall, Englewood Cliffs, NJ.

Rossi, P. E. and Allenby, C. M. (2003). "Bayesian statistics and marketing", *Marketing Science*, 22(3), pp. 304-328.

Rossi, P. E.; McCulloch, R. E. and Allenby, G. M. (1996). "The value of purchase history data in target marketing", *Marketing Science*, 15(4), pp. 321-340.

Rubin, D. B. (1987). *Multiple imputation for non-response in surveys*, Wiley, New York.

Rust, R. T.; Ambler, T.; Carpenter, G. S.; Kumar, V. and Srivastava, R. K. (2004b). "Measuring marketing productivity: current knowledge and future directions", *Journal of Marketing*, 68(4), pp. 76-89.

Rust, R. T. and Chung, T. S. (2006). "Marketing models of service and relationships", *Marketing Science*, 25(6), pp. 560-764.

Rust, R. T.; Lemon, K. N. and Zeithaml, V. A. (2004a). "Return on Marketing: Using Customer Equity to Focus Marketing Strategy", *Journal of Marketing*, 68(1), pp. 109-127.

Rust, R. T., Zeithaml, V. A., and N. Lemon, K. N. (2000). *Driving customer equity; How customer life time value is reshaping corporate strategy*, The Free Press, New York.

Ryals, L. (2002). "Are Your Customers Worth More than Money?", *Journal of Retailing and Consumer Services*, 9(5), pp. 241-251.

Ryals, L. (2005). "Making customer relationship management work: The measurement and profitable management of customer relationships", *Journal of Marketing*, 69(4), pp. 252-261.

Ryals, L. (2006). "Profitable Relationships With Key Customers: How Suppliers Manage Pricing and Customer Risk", *Journal of Strategic Marketing*, 14(2), pp. 101-113.

Ryals, L. and Knox, S. (2007). "Measuring and managing customer relationship risk in business markets", *Industrial Marketing Management*, 36(6), pp. 823-833

Sarel, D. and Marmorstein, H. (2003). "Marketing online banking services: The voice of the customer", *Journal of Financial Services Marketing*, 8(2), pp. 106-118.

Schafer, J. L. and Graham, J. W. (2002). "Missing data: Our view of the state of the art", *Psychological Methods*, 7(2), pp. 147-177.

Schijns, J. M. C. and Schröder, G. J. (1996). “Segment selection by relationship strength”, *Journal of Direct Marketing*, 10(3), pp. 69-79.

Schmittlein, D. C.; Morrison, D. G. and Colombo, R. (1987). “Counting Your Customers: Who They Are and What Will They Do Next?”, *Management Science*, 33(1), pp. 1-24.

Schmittlein, D. C. and Peterson, R. A. (1994). “Customer Base Analysis: An Industrial Purchase Process Application”, *Marketing Science*, 13(1), pp. 41-67.

Sevin, C.H. (1965). *Marketing productivity analysis*, MacGraw-Hill, St. Louis, M.O.

Shah, D.; Rust, R. T.; Parasuraman, A.; Staelin, R. and Day, G. S. (2006). “The path to customer centricity”, *Journal of Service Research*, 9(2), pp. 113-124.

Sharp, A.; Sharp, B. and Wright, M. (2002). “Questioning the value of the “True” brand loyalty distinction”, *International Journal of Research in Marketing*, 19(1), pp. 81-90.

Silvestro, R. and Cross, S. (2000). “Applying the service profit chain in a retail environment: challenging the satisfaction mirror”, *International Journal of Service Industry Management*, 11(3), pp. 244-68.

Singh, S.; Borle, S. and Jain, D.C. (2009). “A generalized framework for estimating customer lifetime value when customer lifetimes are not observed”, *Quantitative Marketing and Economics*, 7(2), pp.181 205

Singh, S. and Jain, D. (2007). “Customer lifetime purchase behavior: An econometric model and empirical analysis”, *Working paper, Rice University, Houston, TX.*

Slater, S. F. and Narver, J. C. (1995). “Market orientation and the learning organization”, Journal of Marketing, 59(3), pp. 63-74.

Smith, M. and Dikolli, S. (1995). “Customer profitability analysis: an activity-based costing approach”, *Managerial Auditing Journal*, 10(7), pp. 3-7.

Spiegelhalter, D. J. (2004). “Incorporating Bayesian Ideas into Health-Care Evaluation”, *Statistical Science*, 19(1), pp. 156-174.

Srivastava, R. K.; Fahley, L. and Christensen, H. K. (2001). “The resource-based view and marketing: The role of market-based assets in gaining competitive advantage”, *Journal of Management*, 27(6), pp. 777-802.

Srivastava, R. K.; Shervani, T. A. and Fahey, L. (1998). “Market-Based Assets and Shareholder Value: A Framework for Analysis”, *Journal of Marketing*, 62(1), pp. 2-18.

Stahl, H. K.; Matzler, K. and Hinterhuber, H. H. (2003). “Linking customer lifetime value with shareholder value”, *Industrial Marketing Management*, 32(4), pp. 267-279.

Storbacka, K. (1994). *The nature of customer relationship profitability: analysis of relationships and customer bases in retail banking*, Doctoral thesis No. 55, Swedish School of Economics and Business Administration, Helsinki.

Storbacka, K. (1997). “Segmentation based on customer profitability – retrospective analysis of retail bank customer bases”, *Journal of Marketing Management*, 13(5), pp. 479-92.

Suh, E. H.; Noh, K. C. and Suh, C. K. (1999). “Customer List Segmentation Using the Combined Response Model”, *Expert Systems with Applications*, 17(2), pp. 89-97.

Swift, R. S. (2001). *Accelerating Customer Relationships: Using CRM and Relationship Technologies*, Prentice Hall, Upper Saddle River, NJ.

The Reporter of Direct Mail Advertising (1941). *How Is Direct Advertising Planned*, April.

The Washington Post (2007). *Shortsighted About the Subprime Disaster*, May, 26th, pp. F02.

Tirenni, G.; Kaiser, C. and Herrman, A. (2007). “Applying decision trees for value-based customer relations management: Predicting airline customers’ future values”, *Database Marketing & Customer Strategy Management*, 14(2), pp. 130–142

Uncles, M. D.; Dowling, G. D. and Hammond, K. (2003). “Customer Loyalty and Customer Loyalty Programs”, *Journal of Consumer Marketing*, 20(4–5), pp. 294–316.

Vakratsas, D. (1998). “Household cost effects on purchase timing decisions: Do demographics matter?”, *Journal of Consumer Marketing*, 15(1), pp. 6-22.

Van Bruggen, G. H. and Wierenga, B. (2005). “When are CRM Systems Successful? The Perspective of the User and of the Organization”. *ERIM Report Series (ERS 2005 048 MKT), Erasmus Research Institute in Management, Rotterdam*.

Van den Poel, D. and Larivière, B. (2004). “Customer attrition analysis for financial services using proportional hazard models”, *European Journal of Operational Research*, 157(1), pp. 196 – 217.

Van Diepen, M. and Franses, P.H. (2006). “Evaluating chi-squared automatic interaction detection”, *Information Systems*, 31(8), pp. 814-831.

Van Raaij, E. M.; Vernooijb, M. J. A. and Van Triest, S. (2003). “The implementation of customer profitability analysis: A case study”, *Industrial Marketing Management*, 32(7), pp. 573-583.

Venkatesan, R. and Kumar, V. (2004). “A customer lifetime value framework for customer selection and resource allocation”, *Journal of Marketing*, 68(4), pp.106-125.

Venkatesan, R.; Kumar, V. and Bohling, T. (2007). “Optimal customer relationship management using Bayesian Decision Theory: An application for customer selection”, *Journal of Marketing Research*, 44(4), pp. 579-594.

Verhoef, P. C. (2004). “The use of econometric models to predict Customer Lifetime Value”, *Medium Econometrische Toepassingen*, 12(3), pp. 22-25.

Verhoef, P. C. and Donkers, B. (2001). “Predicting customer potential value an application in the insurance industry”, *Decision Support Systems*, 32(1), pp. 89-99.

Verhoef, P. C.; Franses, P. H. and Hoekstra, J. C. (2001). “The impact of satisfaction and payment equity on cross-buying: A dynamic model for a multi-service provider”, *Journal of Retailing*, 77(3), pp. 359-378.

Verhoef, P. C.; Franses, P. H. and Hoekstra, J. C. (2002). “The effect of relational constructs on customer referrals and number of services purchased from a multiservice provider: Does age of relationship matter?”, *Journal of the Academy of Marketing Science*, 30(3), pp. 202-216.

Verhoef, P. C. and Lemon, K. (2013). “Successful customer value management: Key lessons and emerging trends”, *European Management Journal*, 31(1), pp. 1-15.

Verhoef, P. C.; van Doorn, J. and Dorotic, M. (2007). “Customer value management: An overview and research agenda”, *Marketing: Journal of Research and Management*, 3(2), pp. 105-120.

Villanueva, J. and Hanssens, D. M. (2007). “Customer Equity: Measurement, Management and Research Opportunities”, *Foundations and Trends in Marketing*, 1(1), pp. 1-95.

Villanueva, J.; Yoo, S. and Hanssens, D. M. (2008). “The impact of marketing-induced versus word-of-mouth customer acquisition on customer equity”, *Journal of Marketing Research*, 45(1), pp. 48-59.

Villas-Boas, M. J. and Winer, R. S. (1999). “Endogeneity in Brand Choice Models”, *Management Science*, 45(10), pp. 1324-1338.

Wang, H. F. and Hong, W. K. (2006). “Managing customer profitability in a competitive market by continuous data mining”, *Industrial Marketing Management*, 35(6), pp. 715-723.

Warner, B. and Misra, M. (1996). "Understanding Neural Networks as Statistical Tools", *The American Statistician*, 50(4), pp. 284-293.

Wayland, R. E. and Cole, P. M. (1997). *Customer connections*, Harvard Business School Press, Cambridge, MA.

Wedel, M. and Kamakura, W.A. (2000). *Market Segmentation: Conceptual and Methodological Foundations*, 2nd Edition, Kluwer Academic Publishers, Norwall, MA.

Weir, K. (2008). "Examining the theoretical influences of customer valuation metrics", *Journal of Marketing Management*, 24 (7-8), pp. 797-824.

Wernerfelt, B. (1984). "A resource-based view of the firm", *Strategic Management Journal*, 5(2), pp. 171-180.

Wiesel, T. and Skiera, B. (2005). "Linking customer metrics to shareholder value for firms with contractual relationships", *Working paper, Institute for the Study of Business Markets, Pennsylvania State University*.

Wiesel, T; Skiera, B. and Villanueva, J. (2008). "Customer Equity: An Integral Part of Financial Reporting", *Journal of Marketing*, 72 (March), pp. 1-14.

Winer, R. S. (2001). "A framework for customer relationship management", *California Management Review*, 43(4), pp. 89- 105.

Wong, J. Y. and Chung, P. H. (2008). "Retaining passenger loyalty through data mining: a case study of Taiwanese airlines", *Transportation Journal*, 47(1), pp. 17-29.

Woo, J. Y.; Bae, S. M. and Park, S. C. (2005). "Visualization method for customer targeting using customer map", *Expert Systems with Applications*, 28(4), pp. 763-772.

Wu, C. and Chen, H.L. (2000), "Counting Your Customers: Compounding Customer's In-Store Decisions, Interpurchase Time, and Repurchasing Behavior", *European Journal of Operational Research*, 127(1), pp. 109-19.

Wu, L.; Liu, L. and Li, J. (2005). "Evaluating customer lifetime value for customer recommendation", *Proceedings for International Conference on Services Systems and Services Management*, 1, pp. 138-143, Institute of Electrical & Electronics Enginee, Chongqing (China).

Wübben, M. and von Wagenheim, F. (2008). "Instant Customer Base Analysis: Managerial Heuristics Often 'Get It Right' ". *Journal of Marketing*, 72(3), pp. 82-93.

Yang, A. X. (2004). "How to develop new approaches to RFM segmentation", *Journal of Targeting, Measurement and Analysis for Marketing*, 13(1), pp. 50-60.

Yoo, S. and Hanssens, D. M. (2005). "Modeling the sales and customer equity effects of the marketing mix", *Working Paper, University of California, Los Angeles, Anderson School of Management.*

Zahavi, J. and Levin, N. (1997). "Applying Neural Computing to Target Marketing", *Journal of Direct Marketing*, 11(1), pp. 5-22.

Zeithaml, V. A.; Rust, R. T. and Lemon, K. N. (2001). "The customer pyramid: Creating and serving profitable customers", *California Management Review*, 42(4), pp. 118-142.

Zhang, J. Q.; Dixit, A. and Friedmann, R. (2010). "Customer loyalty and lifetime value: An empirical investigation of customer packaged goods", *Journal of Marketing Theory and Practice*, 18(2), pp. 127-139.