ISCTE IUL
Instituto Universitário de Lisboa

Department of Quantitative Methods

Modelling Partial Customer Churn in the Portuguese Fixed Telecommunications Industry by Using Survival Models

Sofia Lopes Portela

Thesis submitted as partial requirement for the conferral of

Doctor in Quantitative Methods

Supervisor:
Prof. Doutor Rui Menezes, Associate Professor
ISCTE - IUL

April 2010

ABSTRACT

Considering that profits from customer relationships are the lifeblood of firms (Grant and Schlesinger, 1995), an improvement on the customer management is essential to ensure the competitivity and success of firms. For the last decade, Portuguese customers of fixed telecommunications industry have easily switched the service provider, which has been very damaging for the business performance and, therefore, for the economy.

The main objective of this study is to analyse the partial churn of residential customers in the fixed-telecommunications industry (fixed-telephone and ADSL), by using survival models. Additionally, we intend to test the assumption of constant customer retention rate over time and across customers. Lastly, the effect of satisfaction on partial customer churn is analysed. The models are developed by using large-scale data from an internal database of a Portuguese fixed telecommunications company. The models are estimated with a large number of covariates, which includes customer's basic information, demographics, churn flag, customer historical information about usage, billing, subscription, credit, and other.

Our results show that the variables that influence the partial customer churn are the service usage, mean overall revenues, current debts, the number of overdue bills, payment method, equipment renting, the existence of flat plans and the province of the customer. Portability also affects the probability of churn in fixed-telephone contracts. The results also suggest that the customer retention rate is neither constant over time nor across customers, for both types of contracts. Lastly, it seems that satisfaction does not influence the cancellation of both types of contracts.

Keywords:

Survival models, frailty models, customer churn, customer management.

Classification:

C41, M31

RESUMO

Considerando que os lucros gerados pelos clientes são vitais para as empresas (Grant e Schlesinger, 1995), uma melhoria na gestão do cliente é fundamental para assegurar a competitividade e o sucesso das empresas. Na última década, os clientes portugueses das empresas de telecomunicações fixas têm mudado de operador com demasiada facilidade, o que tem prejudicado o desempenho das empresas e, consequentemente, a economia.

O principal objectivo deste estudo é analisar o cancelamento de contratos de telefone fixo e ADSL por clientes residenciais, através do uso de modelos de sobrevivência. Para além disso, pretende-se testar o pressuposto de que a taxa de retenção de clientes é constante ao longo do tempo e entre clientes. Por último, pretende-se analisar o efeito da satisfação do cliente no cancelamento destes tipos de contratos. Os modelos são construídos com base numa base de dados de larga escala fornecida por uma empresa portuguesa deste sector. Os modelos são estimados com base num vasto número de variáveis, incluindo informação básica sobre o cliente, dados demográficos, indicação sobre o cancelamento do contrato, dados históricos sobre o uso dos serviços, facturação, contracto, crédito, etc..

Os resultados mostram que as variáveis que influenciam o cancelamento de ambos os tipos de contratos são o uso do serviço, a facturação média, o valor em dívida, o número de facturas em dívida, o método de pagamento, o método de pagamento do equipamento, a existência de tarifas planas e o distrito do cliente. A portabilidade de número parece influenciar o cancelamento de contratos de telefone fixo. Os resultados também mostram que a taxa de retenção de clientes não é constante ao longo do tempo nem entre clientes em ambos os tipos de contratos. Por último, parece que a satisfação não influencia o cancelamento de ambos os tipos de contratos.

Palavras-chave:

Modelos de sobrevivência, modelos de heterogeneidade não observada, *churn*, gestão do cliente.

Classificação:

C41, M31

EXECUTIVE SUMMARY

Considering that profits from customer relationships are the lifeblood of firms (Grant and Schlesinger, 1995), an improvement on the customer management is essential to ensure the competitivity and success of firms, mainly in a period of economic recession. For the last decade, Portuguese residential customers of fixed telecommunications industry have easily switched the service provider, mainly due to the strong competition on this industry and the low switching costs, which has been very damaging for the business performance and, therefore, for the economy. As such, researchers have recognised the importance of an in-depth study of customer churn (*i.e.*, the customer's decision to terminate the relationship with the provider). In this context, an *a priori* knowledge about the risk of a given customer to cancel a given contract with the service provider at any time is a valuable tool that allows firms to take preventive measures to avoid the defection of potentially profitable contracts from customers.

The main objective of this study is to analyse the partial churn of residential customers in the fixed-telecommunications industry (fixed-telephone and ADSL), by using survival models. Additionally, we intend to test the common assumption of constant customer retention rate over time and across customers. Lastly, the effect of satisfaction on partial customer churn is analysed.

The models are developed by using large-scale data from an internal database of a Portuguese fixed telecommunications company. The models are estimated with a large number of covariates, which includes customer's basic information, demographics, churn flag, customer historical information about usage, billing, subscription, credit, and other.

Our results show that customers with harder usage of the fixed-telephone service have a longer relationship with the service provider. As regards to the ADSL contracts, the results provide evidence that the probability of churn does not vary with the internet usage, but customers with more additional usage than those contracted have longer relationships with the service provider. Moreover, it seems that both types of contracts with flat plans have a lower risk of churn than those without flat plans. The results of this study also indicate that customers with greater average monthly spending with the service provider have shorter

contract lifetimes of both types. Moreover, it seems that the total number of overdue bills (since ever) negatively affect the survival time of both kind of contracts in study. It also seems that the survival time of fixed-telephone contracts of customers that required portability is larger than the one that did not require portability. Contracts paid by direct debit also last longer than contracts paid by other methods. Furthermore, the contracts of those customers who buy the necessary equipment last longer than those of customers who rent the equipment. The results of the model appear to indicate that the probability of churn varies across some provinces.

The results also suggest that the customer retention rate is neither constant over time nor across customers, for both types of contracts.

Lastly, it seems that satisfaction does not influence the cancellation of both types of contracts.

ACKNOWLEDGMENTS

During the course of this research, many people and entities have provided me support, and it is a pleasure to express my acknowledgments to all of them.

Many thanks go first and foremost to my supervisor, Prof. Rui Menezes, who was always there to give advice, suggestions, encouragement, and friendship. I would like to thank him for all the time and energy he has dedicated to the supervision of this dissertation.

Special thanks to the ISCTE-IUL who funded my research and granted me a leave of absence.

I would like to thank my colleagues João Pedro Pereira, João Pedro Nunes, and Clara Raposo for their interest in discussing some topics in finance. I also would like to thank my colleague and friend José Farinha for all the support on the use of the database manipulation software.

I thank Ricardo Reiçadas from the library of ISCTE-IUL for his permanent support on gathering books and papers from other universities.

I wish to thank the firm that provided their internal customer database, without which this study would be not possible, and which strongly believes on the importance of the expected results to the improvement of its customer portfolio management and, therefore, to the improvement of its profitability.

I am grateful to acknowledge the Fundação para a Ciência e Tecnologia for their financial support (Project reference: PTDC/GES/73418/2006), who allows the presentation of some results of this research in some international conferences.

I would specially like to thank my friends Manuela and João Neves for their emotional support in good and bad periods of my life.

Finally, I want to thank all my family for their love, encouragement, and unconditional support. They teach me to always work to reach my objectives and I am very grateful for that. Papy, wherever you are, I know you are with me always and forever. Lastly, and specially, thanks to my boyfriend, for his love, encouragement, assistance, and patience for all the time I spent away from him.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

The Portuguese market of fixed telecommunications soared in the last decade and, as a consequence, firms have focused on customer acquisition and neglected customer retention[1]. This period is characterized by a strong business competition and low switching costs in this industry, which gave rise to a phenomenon of customer defection, and, thus, high customer churn rates, which has serious consequences for the business financial performance and, therefore, for the economy. According to several researchers, customer churn (*i.e.*, the customer's decision to terminate the relationship with a provider) is the main reason of profitability losses in the telecommunications industry (TI henceforth), due to losses on current and potential revenues, marketing costs, and brand image (*e.g.*, Ahn *et al.*, 2006; Nath and Behara, 2003; Qian *et al.*, 2006; Seo *et al.*, 2007; Zhang *et al.*, 2006).

Nevertheless, the fixed telecommunications market is becoming saturated in Portugal and, as consequence, the pool of "available customers" is limited and, as such, firms need to change their strategy from customer acquisition to the retention of potentially valuable customers (Hadden *et al.*, 2005; Hung *et al.*, 2006), because firms cannot lose valuable customers to their competitors. Bolton and Tarasi (2006) suggest that customer retention is often easier and cheaper than customer acquisition in stable markets with low growth rates.

The customer retention became a buzzword in the 1990s, mainly due to the work of Reichheld and Sasser (1990), who firstly provided evidence about the advantages of customer retention. Although their results definitively caused a change in the marketing theory, they are not consensual (see, for example, Carroll, 1991/92; Dowling and Uncles, 1997; East *et al.*, 2006; Gupta *et al.*, 2006; Ranaweera, 2007; Reinartz and Kumar, 2000). Following this new paradigm, many firms have focused on retaining all customers. Nevertheless, many researchers argue that the retention strategy must be strongly linked with the customer lifetime value (*i.e.*, the expected net present value of the future cash flows of the customer – CLV), and, consequently, enterprises should not try to retain all of their current customers, because they are probably investing in unprofitable customers (Gupta and Lehmann, 2003; Jain and Singh, 2002; Malthouse and Blattberg, 2004; Ryals,

[1] By fixed telecommunications industry we mean firms that provide fixed-telephone, internet, and pay-TV.

2003; Thomas *et al.*, 2004), and, in this way, they are destroying value (Gupta and Lehmann, 2005; Jain and Singh, 2002; Ryals, 2003).

Researchers have recognised the importance of an in-depth study of customer churn. The customer churn issue is present both in studies about CLV as a component of CLV and on specific studies of churn, but in different perspectives. In studies about CLV, customer churn is mainly analysed in a theoretical way, whereas on the later case, the statistical models with empirical data are predominant. Customer churn has been studied using different techniques, in different industries (*e.g.*, banking, insurance, telecommunications), and in different contexts (contractual vs. noncontractual settings, continuous vs. discrete time). Ahn *et al.* (2006) point out that the reasons of customer churn and the customer behaviour towards churn need to be more studied.

Despite the large amount of research done on customer churn, there are only few studies applied to the fixed telecommunications industry. Furthermore, to the best of our knowledge, none is applied to firms that simultaneously offer ADSL, fixed line telephone, and pay-TV. The majority of published research about customer churn prediction in the telecommunications industry analyses the mobile telecommunications. This issue has never been studied in Portugal. Many studies focus on model accuracy or comparison of techniques rather than on testing the effect of churn covariates. Many of these studies were presented in conferences about data-mining; so the majority of them apply data-mining techniques. Lastly, most of them model whether (or not) a customer is likely to churn in a pre-specified time period, rather than the longitudinal churn pattern over the duration of the relationship.

It seems relevant to do a more detailed study of the customer churn in the fixed telecommunications industry, because it may be misleading to make decisions based only on the results of the mobile telecommunications industry, which presents very different characteristics. Moreover, considering that the customer churn behaviour may be influenced by the customer culture, it is pertinent to examine different markets, like the Portuguese one. In this context, the aim of this study is to analyse the partial churn of residential customers in the fixed-telecommunications industry (fixed-telephone and ADSL) in Portugal, by using continuous survival models. It also intends to analyse the assumptions of constant retention rate over time and across customers. Lastly, the effect of satisfaction on partial customer churn is analysed. The models are developed by using large-scale data from an internal database of a Portuguese company which presents offers of fixed telephone, ADSL, and pay-TV.

Some of the specific areas where these models can help the customer management are: (i) *a priori* knowledge about the probability (risk) of a given customer to cancel a contract with the service provider at any time and, thus, firms can take preventive measures to avoid the defection of potentially profitable contracts of customers, (ii) customer selection to retention programs; (iii) marketing resource allocation across customers; and (iv) computation of CLV.

The study is organized as follows. Chapter 2 reviews the theoretical and empirical research that has been developed in the customer management context, and more specifically, in the customer churn context. Chapter 3 is concerned with continuous survival models. Chapter 4 presents the data and the empirical results of the study. Finally, Chapter 5 summarizes the main results, presents the data limitations of the study, and suggests directions for further research.

2. LITERATURE REVIEW

2.1. Introduction

In this chapter we present a review of literature about the main current areas of interest in the customer portfolio management that are related to customer churn. As such, section 2.2 introduces the notion of customers as assets that has emerged in the literature. In section 2.3 we describe the types of relationships with customers and opportunities for transactions. In section 2.4 we summarize the evolution of the main principles of customer portfolio management; a review of the desirable customer portfolio is presented, and, lastly, the notion of customer equity. Section 2.5 reviews the notion of customer lifetime value and some proposed models to compute the value of a customer. Lastly, section 2.6 presents a review of literature about customer churn, namely its notion, some studies about the prediction of customer churn in the telecommunications industry, the variables used in the customer churn prediction, and some available models to estimate customer churn.

2.2. Customers as (intangible) assets

Although the customer-centric paradigm is more than 50 years old (Shah *et al.,* 2006), the customer revolution only happened in the 80s (Boyce, 2000), during which the firms were encouraged to focus on customers rather than on products. In this way, customers became the centre of the organizations (Boyce, 2000). For a long time, researchers suggested that the "customer is always right" and thus managers focused in satisfying the customer needs and improving the customer satisfaction. This period is referred by Gupta and Lehmann (2005) as the "traditional marketing strategy". These authors argue that a new paradigm has emerged, which they denominate "customer-based strategy", and which is also designated "customer equity management" by Blattberg *et al.* (2001) and Hogan *et al.* (2002b). The main difference between these paradigms is that the traditional marketing strategy was only concerned with the value that a firm provides to a customer, and the customer-based strategy is also concerned with the value of a customer to the firm. Thus, this approach emphasises the two sides of customer value. According to the customer-based approach, the firm should invest to provide value to the customer and, in counterpart, the customer should provide returns to the firm and its shareholders (Bolton and Tarasi, 2006). Boyce (2000) and Johnson and Selnes (2005) do suggest that the marketing thinking reveals an evolution from the "customer is king" to the "customer is cash".

In a general way, researchers argue nowadays that customers should be viewed as assets of firms (*e.g.*, Anderson *et al.*, 1994; Bell *et al.*, 2002; Berger *et al.*, 2002; Blattberg *et al.*, 2001; Boulding *et al.*, 2005; Colombo and Jiang, 1999; Dhar and Glazer, 2003; Gupta and Lehmann, 2003, 2005; Gupta *et al.*, 2004; Hogan *et al.*, 2002a, 2002b; Kumar *et al.*, 2006; Ryals, 2002a; Stahl *et al.* 2003; Wayland and Cole, 1994; Weinstein, 2002; Woo *et al.* 2005; Wyner, 1996). This new understanding of the role of customers in the firm was firstly adopted by relationship marketing researchers, mainly in the business-to-business domain (Hogan *et al.*, 2002b). Even though the treatment of customers as assets has been widely discussed, the value of customers to the firm has been examined with excessive superficiality and little rigour. Furthermore, there are relatively few rigorous empirical studies about this issue.

Some researchers argue that customers are intangible assets of firms (*e.g.*, Dhar and Glazer, 2003; Forbes, 2007; Gupta and Lehmann, 2003; Kumar and George, 2007; Ryals, 2002a, 2002b) because customers are not "owned" by the firms. In fact, firms only have a relationship with them, and even this relationship might be not exclusive (Dhar and Glazer, 2003).

2.3. Type of relationships with customers and opportunities for transactions

The type of relationship a firm has with its customers was firstly categorized by Jackson (1985) into two main groups, which are:

- *Lost-for-good* – also called customer retention situation, contractual setting, and subscription business setting by Dwyer (1989), Reinartz and Kumar (2000), and Fader and Hardie (2006), respectively (henceforth, contractual setting); and

- *Always-a-share* – also called customer migration situation, noncontractual setting, and visitant business setting by Dwyer (1989), Reinartz and Kumar (2000), and Fader and Hardie (2006), respectively (henceforth, noncontractual setting).

The main characteristics of these two types of relationships with customers are presented in Table 1.

Although this classification is widely accepted in the customer value literature, Singh (2003) criticises it arguing that these two groups are not mutually exclusive. In fact, a customer who cancels the relationship with a provider in a contractual setting may reinitiate a new relationship with the firm, *i.e.,* he/she is not lost forever. On the other hand, a customer who terminates the relationship with a provider in a noncontractual setting may never come back, *i.e.,* he/she is not always-a-share, but lost-for-good.

Customers have two types of opportunities for transactions, which are: (i) continuous time; and (ii) discrete time. In the former case, transactions can take place at any point in time (examples: supermarkets, credit card, doctor visits, hotel stays) and in the later case, transactions can only occur at fixed points in time (examples: magazine subscriptions, apartment rental).

Table 1– Characteristics of the two types of relationships with customers

Contractual setting	Noncontractual setting
There is a contract with the customer or the customer signed a subscription.	There is not any contract with the customer nor a subscription.
Defection is observable because: (i) the customer needs to contact the firm to cancel his/her contract; or (ii) the customer fails to renew his/her contract/ subscription. (Fader *et al.*, 2005b, 2006; Fader and Hardie, 2006, 2006b; Gupta and Zeithaml, 2006)	Defection is not directly observable because the customer does not explicitly terminate the relationship, as the customer does not notify the firm when he/she becomes inactive to the firm.. (Fader *et al.*, 2004, 2005b, 2006; Gupta and Zeithaml, 2006; Reinartz and Kumar, 2003)
The customer buys a given product/service from an unique supplier.	The customer typically splits his/her expenses on a specific produtct/service among several firms. (Bolton, 1998; Gupta and Zeithaml, 2006; Reinartz and Kumar, 2000; Rust *et al.*, 2004; Singh, 2003; Stahl *et al.*, 2003)
The customer is either totally committed to the provider or totally lost, *i.e.*, customer defection is permanent (once a customer leaves the firm, it is assumed that he/she does not return). So, if an “ex-customer” purchases again in a future moment, he/she will be treated as a new customer. (Bauer *et al.*, 2003; Bell *et al.*, 2002; Calciu and Salerno, 2002; Dwyer, 1989; Gupta and Zeithaml, 2006; Rust *et al.*, 2004; Singh, 2003; Stahl *et al.*, 2003; Venkatesan and Kumar, 2004)	The customer is never lost forever. The customer can make a purchase from a firm, leave the firm, and either purchase from a competitor or not purchase at all in a period and then purchase again. Therefore, firms cannot differentiate a customer who has terminated his/her relationship with them from a customer who is the middle of a break between transactions. (Calciu and Salerno, 2002; Fader *et al.*, 2004; Reinartz and Kumar, 2003; Singh, 2003)
Managers can predict the probability of customer retention and customer defection based on historical data. (Bauer *et al.*, 2003; Fader and Hardie, 2006; Rust *et al.*, 2004; Schweidel *et al.*, 2006; Stahl *et al.*, 2003)	As defection time is not observed, neither the notion of “retention rate” nor survival analysis cannot be meaningfully used. Thus, the focus is on inferring if a customer is still “active” and on predicting future activity. (Calciu and Salerno, 2002; Fader and Hardie, 2006; Gupta and Zeithaml, 2006; Venkatesan and Kumar, 2004)
Managers can accurately predict the customers revenues, based on the customer usage. (Bolton, 1998; Reinartz and Kumar, 2000)	It is more difficult to make predictions of the customer’s revenues in the long-run.
Examples of industries: utilities, mobile phones (post-paid), ISPs, credit card, magazine subscriptions	*Examples of industries:* grocery stores, doctor visits, hotel stays, supermarkets, mobile phones (pre-paid)

2.4. Customer Portfolio

2.4.1. Evolution of the main principles of customer portfolio management

The customer management has been oriented towards different principles over time. For a long time, managers focused on firm growth, and, thus, on customer acquisition (Rosenberg and Czepiel, 1983). More recently, a new paradigm has been suggested by researchers, which is based on customer retention.

Even though Reichheld and Sasser (1990) were not pioneers in pointing out the advantages of customer retention (East *et al.*, 2006), they firstly provided evidence about those advantages, which are based on a strong relationship between customer retention and profitability. They found that long-time customers (i) spend more over time, (ii) the operating costs to serve them decline over time, (iii) become more loyal and then promote the word-of-mouth, and (iv) are less price-sensitive. Furthermore, Reichheld and Kenny (1990) point out that the expense of acquiring a new customer occurs only once and at the beginning of the relationship.

Additionally, Reichheld (1996) argues that customer defection has severe effects on firms' profitability because firms have to incur in heavy costs to acquire new customers and older customers usually generates greater cash flows and profits than newer ones. It is popularly believed that the acquisition of a new customer costs at least five times more than keeping an existing one. These conclusions caused a change in the marketing theory, since researchers started definitely arguing that enterprises should focus more on customer retention rather than on customer acquisition (*e.g.,* Reichheld, 1996; Thomas *et al.*, 2004; Trubik and Smith, 2000; Weinstein, 2002). Blattberg and Deighton (1996) emphasise that firms should decide the balance between customer acquisition and retention investment according to the industry and the customer behaviour, because the concept of customer retention is difficult to implement in certain industries. Thakur and Summey (2005) argue that the existence of long lasting relationships with customers is crucial to the rise of market share and long-term competitive advantages of firms.

Although many people agree with the arguments of Reichheld and Sasser and started following their paradigm (*e.g.,* Berger and Nasr, 1998; Colgate *et al.*, 1996; Hansotia, 2004; Kim *et al.*, 2005; Seo *et al.*, 2007; Weinstein, 2002), some authors have questioned them. For instance, Carroll (1991/92) heavily criticises Reichheld and Sasser. She defends that retail bank customers do not get more profitable with tenure. Dowling and Uncles (1997) argue that, considering the heavy costs incurred to retain customers, it is not clear that long-time customers are less expensive to serve. Reinartz and Kumar (2000) studied empirically the Reichheld and Sasser's propositions in a large catalog retailer (noncontractual setting) and they found that long-life customers are not necessarily profitable customers. Also Reinartz and Kumar (2002: 87) found that "there is little or no evidence to suggest that customers who purchase steadily from a company over time are necessarily cheaper to serve, less price sensitive, or particularly effective at bringing in new business". Jain and Singh (2002) argue that the propositions of Reichheld and Sasser have not been carefully tested. East *et al.* (2006) present a review of the Reichheld and Sasser's propositions and they concluded that (i) the evidence provided by the authors is erratic and often weaker than suggested; (ii) the potential financial gains from customer acquisition can be much larger than gains via defection reduction; and (iii) much of the defection is near-involuntary. Gupta *et al.* (2006) mention that managers may believe that they spend more on customer acquisition than customer retention because the customer acquisition costs are easily quantified, while customer retention costs are not. In a study about the effect of customer satisfaction and the duration of the relationship on several variables, in the fixed line telephone industry in South-Eastern England, Ranaweera (2007) did not find support for all of the Reichheld's propositions. He found that the duration of the relationship is positively associated with the spending amount and negatively associated with positive and negative word-of-mouth. He did not find support for the relationship between duration of the relationship and price sensitivity. Moreover, he provides evidence that highly satisfied customers who have a long relationship with the service provider are more likely to be less price sensitive and are less likely to give positive WOM (which contradicts the theory). Shapiro *et al.* (1987) argue that, over time, customers are likely to become more price sensitive through rival product offerings, which are often at lower prices. Villanueva and Hanssens (2007) present some arguments in favour and against the Reichheld and Sasser' propositions.

Reichheld and Sasser (1990) defend that firms should develop a customer management toward zero defections. Nevertheless, Blattberg *et al.* (2001) classify this idea as a myth seeing that there are some exogenous and uncontrollable factors (*e.g.,* customers' desire for newness) that affect the customer retention potential and, thus, a maximum 100 percent retention rate is not possible. Moreover, they also emphasise that retention is not free. Van den Poel and Larivière (2004) point out that a retention rate of 100% is utopian, due to uncontrollable reasons of defection, like, for example, natural death or moving to a foreign country. Gupta and Lehmann (2005) make clear that a rate of retention of 100 percent would only be possible if the firm had few customers and those customers were either extremely loyal or had no alternative except to stay loyal.

Simultaneously, several researchers argue that firms should neither focus on nor try to retain all of their current customers, because they are probably investing in unprofitable customers (Gupta and Lehmann, 2003; Jain and Singh, 2002; Malthouse and Blattberg, 2004; Reichheld 1991/92; Ryals, 2003a; Thomas *et al.*, 2004), and, in this way, they are destroying value (Gupta and Lehmann, 2005; Jain and Singh, 2002; Payne *et al.*, 2000; Ryals, 2003a; Wayland and Cole, 1994) because (i) the retention of unprofitable customer is damaging to the firm, and (ii) the money wasted on the retention of unprofitable is not used on the retention of profitable ones, who are harder to get (Thomas *et al.*, 2004).

In conclusion, the retention strategy should be strongly linked with the customer value (Blattberg and Deighton, 1996; Payne *et al.*, 2000).

2.4.2. The desirable customer portfolio

Nowadays, customer portfolio management is a very important discipline because profits from customer relationships are the lifeblood of firms (Grant and Schlesinger, 1995; Gupta and Lehmann, 2005). According to the last developments, the main input to customer portfolio management is the customer value.

It has been suggested that managers should invest in the retention and acquisition of potentially profitable customers and reduce or cease relationships with those customers that probably cannot become profitable (Bolton and Tarasi, 2006; Peppers and Rogers, 2004; Ryals, 2002a, 2003a; Thakur and Summey, 2005; Woo and Fock, 2004; Zeithaml *et al.*, 2001). Researchers argue that the abandonment of negative value customers allows the re-allocation of resources to the positive value customers, which will have more resources available. Peppers and Rogers (2004) mention that before firing customers, firms should evaluate the chance to convert the unprofitable customers into profitable ones with incentives.

Nevertheless, Nasr-Bechwati and Eshghi (2005) argue that fired customers can propagate bad word-of-mouth about the firm, which could be very damaging because it can affect the acquisition and the retention capacity of the firm and, moreover, the fired customers will be very difficult to win back. Mittal *et al.* (2008) also present some risks of customer divestment. They emphasise that customer divestment may cause: (i) the lost of valuable sources of information, experimentation, and innovation; (ii) the changing of the competitive dynamics due to the accommodation of customers by a rival firm; (iii) insecurity on the remaining customers because they may wonder they are next; (iv) the increase of costs in the remaining customers in firms with high fixed costs; and (v) downsized customer base. Peppers and Rogers (2004) argue that firing unprofitable customers is not a hostile activity; instead, they argue that it allows the fair distribution of value. Bolton and Tarasi (2006) suggest that instead of firing customers, firms can offer a less attractive value proposition to these customers (*e.g.*, high prices or low-quality products).

Many researchers have argued that firms should focus on their most profitable customers (*e.g.*, Blattberg and Lehmann, 2005; Duboff, 1992; Malthouse and Blattberg, 2004; Nasr-Bechwati and Eshghi, 2005; Ryals, 2003a) or on their most valuable customers (*e.g.*, Jain and Singh, 2002; Malthouse and Blattberg, 2004; Mulhern, 1999; Payne and Frow, 1999; Peppers and Rogers, 2004; Reichheld and Sasser, 1990; Weinstein, 2002; Wyner, 1996). Wyner (1996) proposes that fewer resources should be allocated to lower-value customers. Nevertheless, some researchers suggest that firms should not invest exclusively on the current profitable customers because, in this way, few resources will remain to be used in the attraction of the current less profitable customers with high potential value through up-selling activities (Nasr-Bechwati and Eshghi, 2005). Reichheld (1993) suggests that firms should target those customers who are likely to buy products or services to the firm over time, instead of those who are easiest to attract or the most profitable in the short term. Peppers and Rogers (2004) argue that the main aim of the customer managers should be to maximize the long-term value of his customer portfolio by keeping and growing his customers. Other researchers argue that (i) focusing on a limited number of customers may ignore the possible economies of scale and (ii) lower-priority customers can become dissatisfied and then they might defect or diffuse negative WOM (Homburg *et al.* 2008). Woo *et al.* (2005) suggest that firms should find and keep the profitable or potentially profitable customers. Johnson and Selnes (2004, 2005) advocate that customer portfolio should include both closer and weaker customer relationships with the firm and customers who have steady and volatile purchasing patterns.

On the other hand, some researchers claim for the use of the Pareto Analysis, which indicates that 20 percent of the customers with the highest sales volume generate 80 percent of profits; thus, they argue that firms should focus on these 20 percent most profitable customers. For instance, Malthouse and Blattberg (2004) suggest that firms should focus on the 20 percent of customers with greater lifetime value because they are the best customers. McClymont and Jocumsen (2003) also mention that the “right” customers are those 20 percent of loyal, highly profitable. Thakur and Summey (2005) use the Pareto principle to classify the customer portfolio into profitable customers, potentially profitable customers and not-so-profitable customers. Nevertheless, this measure may be misleading because customers with the highest sales volume may not be the most profitable (Ang and Taylor, 2005; Stahl *et al.*, 2003).

As presented in section 2.2, researchers recognise that customers are assets of firms. Consequently, customers should be evaluated like any other asset, and, thus, the customer portfolio can be managed like any other asset’s portfolio. Portfolio management emerged in the 1960s as a result of the work of Markowitz in financial markets. Later, portfolio management expanded to other areas, like strategic or product management (Turnbull, 1990). The modern portfolio theory proposes an efficient management of the financial portfolios, *i.e.*, a better allocation of the limited resources, which will maximise the return for a given level of risk or minimise the risk for a given return. Markowitz proposes that investors should have a diversified portfolio, which should include both high-risk, high-return and low-risk, low-return assets, which means that the objective of investors is not profit maximisation, because it can lead to an undesirable level of risk. According to him, both returns and risk have to be taken into consideration in order to make correct investment decisions. Some researchers suggest the application of these principles to the customer portfolio management (*e.g.*, Bolton *et al.*, 2004; Bolton and Tarasi, 2005; Dhar and Glazer, 2003; Rust *et al.*, 2004; Ryals, 2002b, 2003b; van Amelsfort, 2006; and von Wangenheim and Lentz, 2005). The application of the financial portfolio theory to the marketing context has been largely criticised by Devenney and Stewart (1988). They argue that the main assumptions of the modern portfolio theory are not satisfied in product market context. Some authors also argue that Modern Portfolio Theory is computationally complex. They did not find support to the idea that there are groups of customers that have a systematically negative beta or that, in other words, are good "risk-reducers" for the customer portfolio.

Apparently, the philosophy of focus on the most profitable customers does not consider entirely the Portfolio Theory of Markowitz, because it only focuses on value and ignores the risk. Furthermore, the cash flows generated by a portfolio of customers are almost always less volatile than the cash flows of any customer individually (Dhar and Glazer, 2003). Also, the strategy of focusing only on a given cluster may be very risky because firms may become dependent on it.

One of the preeminent questions on customer portfolio management is on what customers to invest (and how much) and on what customers to disinvest. The majority of firms make these decisions based only on intuitive rules. Nevertheless, it seems that optimization models can make a great contribution to these decisions. Bonfrer *et al.* (2007) also consider customer portfolio optimization a promising area of study. Bolton and Tarasi (2006:.18) do argue that "decisions about individual customers cannot be made without considering the optimal characteristics of the entire customer portfolio". Dhar and Glazer (2003) argue that firms should decide which customers to acquire or retain based on the effect that a specific customer will have on risk and return of the customer portfolio.

Rosenberg and Czepiel (1983) and Blattberg *et al.* (2001) suggest that an optimal customer portfolio is gathered from a combination of new and repeat buyers, which is the result of acquisition and retention spending. Mondschein and Musalem (2004) present an extensive review of literature about the optimal resource allocation across customer segments and they also propose some models of optimal resource allocation to customer relationship management.

2.4.3. Customer Equity: The value of the customer portfolio

The concept of customer equity (CE henceforth) was firstly introduced by Blattberg and Deighton (1996) (Blattberg *et al.*, 2001; Drèze and Bonfrer, 2005; Hogan *et al.*, 2002b; Rust *et al.*, 2000). According to them, CE is the "discounted, expected contributions of all current customers" (p. 138), that is, CE is the present value of the current customer portfolio of the firm. Since then, this concept has been largely studied.

The definition of CE is not consensual. Some researchers argue that CE is the individual customer lifetime value summed over all current and future customers (Bauer and Hammerschmidt, 2005; Bauer *et al.*, 2003; Bayón *et al.*, 2002; Gupta and Lehmann, 2005; Gupta *et al.*, 2004; Hansotia, 2004; Hogan *et al.*, 2002a; Rust *et al.*, 2004; Villanueva and Hanssens, 2007), while others suggest that CE is the sum of the lifetime value of all the firm's customers (Gupta *et al.*, 2006; Kumar *et al.*, 2004; Lemon *et al.*, 2001; Niraj *et al.*, 2001; Rust *et al.*, 2000). But the later definition can have several interpretations, because "all" can be interpreted as all current customers or all current and potential customers. For example, Rust *et al.* (2000) and van Wangenheim and Lentz (2005) clarify that they include only the current customers, and Gupta *et al.* (2006) include the current and future customers, but many other researchers do not clarify this point. Bayón *et al.* (2002) present some problems that may emerge from the inclusion of the potential customers on CE.

2.5. Customer value: The CLV

Considering customers as assets, some authors point out that it is crucial to calculate their financial value to the firm (*e.g.*, Boyce, 2000; Drew *et al.*, 2001; Gupta and Lehmann, 2003; Jain and Singh, 2002; Malthouse and Blattberg, 2004; Pfeifer *et al.*, 2005; Reichheld, 1996; Wyner, 1996).

The idea of valuing customers arose some decades ago, much earlier than the spread of the relationship marketing theory. For instance, Sevin (1965) (cited in Bell *et al.*, 2002) design a methodology for the calculation of the individual customer profitability. Kotler (1974) suggests the valuation of the long-run customer profitability. Nevertheless, only since the end of the 1980s customer valuation theory and practice has soared (Donkers *et al.*, 2003). The first applications of customer valuation were in direct marketing (Villanueva, 2003; Villanueva and Hanssens, 2007). Although the customer value measurement presents a great development (Bell *et al.*, 2002), this concept has not been widely applied, due to the necessity of enormous amounts of data and sophisticated models (Gupta and Lehmann, 2003). Furthermore, being intangible assets, customers are difficult to evaluate with precision (Gupta and Lehmann, 2003). Hogan *et al.* (2002a) emphasise that customer value models are still in the infancy stage.

Generally, researchers proposed that the value of a customer is the expected net present value of his/her cash flow stream. In this way, the customer value is an application of the principles of contemporary finance to evaluate customers, more precisely the discounted cash flow method (Day and Fahey, 1988; Drèze and Bonfrer, 2005). This concept was proposed by Rappaport in 1986 and became popular in corporate valuation. The customer value is usually called customer lifetime value and is often abbreviated CLV or LTV (henceforth CLV). Other authors have used different names to denote CLV. For instance, Jain and Singh (2002) and Mulhern (1999) adopt customer profitability, Berger and Nasr (1998) adopt the term economic worth of a customer, and Pfeifer and Farris (2004) designate it as expected customer future value. Nowadays, CLV is the most popular customer measure because it is forward-looking, includes all the elements of customer profitability and it is an essential element of the customer-centric paradigm (Kumar and Shah, 2004). In fact, CLV has become a buzzword in the last years (Nasr-Bechwati and Eshghi, 2005). CLV may be a useful measure helping the decision making, both on operational and strategic marketing decisions and even on strategic decisions of the firm (*e.g.*, customer segmentation, customer selection, marketing resource allocation across customers, guidance for marketing investments, customer base valuation, firm valuation, etc.).

CLV is a more powerful measure than historic customer profitability analysis, because CLV looks at the future potential of the customer, whereas current and past profitability is not forward-looking (Boyce, 2000; Jain and Singh, 2002; Reinartz and Kumar, 2000; Ryals, 2002a). Customer profitability is the difference between revenues and costs associated with the customer during a specific period of time (Boyce, 2000; Pfeifer *et al.*, 2005) and this measure is calculated on a single period basis, usually the last economic year (Ryals, 2006). In this way, unlike CLV, customer profitability is not a good basis for developing marketing strategies (Ryals, 2002a).

The process of CLV calculation should take into consideration the cash flow patterns over time (Nasr-Bechwati and Eshghi, 2005), the relationship birth, purchase activity, and the defection (Reinartz and Kumar, 2000).

CLV has been widely studied and, as a result, a huge number of models is available in the literature. The sophistication of the models varies a lot, since simple models to more complex ones (which aim to incorporate the complexities of the real business situations). Several researchers have intended to evaluate the customers, estimating their lifetime value (*e.g.*, Berger *et al.*, 2003; Drew *et al.* 2001; Fader *et al.*, 2005a; Gupta and Lehmann, 2003, 2005; Venkatesan and Kumar, 2004), but the majority of them only proposed formulas to evaluate the customer value (*e.g.*, Berger and Nasr, 1998; Gurau and Ranchhold, 2002; Peppers and Rogers, 2005; Pfeifer and Farris, 2004; Pfeifer *et al.*, 2005). Most researchers do not propose methods to forecast the CLV components.

There is not a generally accepted customer value calculation formula. Nevertheless, the majority of proposals are based on one of the following formulas:

$$CLV = \sum_{t=1}^{T} \frac{revenues_t - \cos ts_t}{(1+i)^t}$$

and

$$CLV = \sum_{t=1}^{T} \frac{cash\ flow_t}{(1+i)^t}$$

Table 2 presents a summary of the literature review on CLV computation.

Table 2 - Summary of the literature review on customer lifetime value computation

Authors	CLV formula [2]
Anderson *et al.* (1994)	$CLV = \sum_{t=1}^{T} \lambda \, m'' \left(\frac{\Pr\{Loyal \mid Satisfaction\}}{1+i} \right)^{t/\lambda}$
Blattberg and Deighton (1996)	$CLV = a\, m' - A + a \times \left(m' - \frac{c'}{r} \right) \times \left[\left(\frac{r}{1+i} \right) \Big/ \left(1 - \frac{r}{1+i} \right) \right]$
Gloy *et al.* (1997)	$CLV = \sum_{t=1}^{T''''} \left\{ (1-a')^{t-1}\, a' \left[\begin{array}{l} -\frac{A_t}{(1+i)^{t-1}} - \frac{A_t'' \times n''_t}{(1+i)^{t-1}} - \frac{A_t'''}{(1+i)^{t-1}} - \sum_{j=0}^{t} \frac{W_t' \times n_t'''}{(1+i)^j} \\ + \sum_{j=0}^{T''''-t} \frac{r^j \left[\left(R''_{j+t} - C''''_{j+t} \right) \times n''_{j+t} - C'''''_{j+t} \times n''_{j+t} - C''''''_{j+t} \right]}{(1+i)^{j+t}} \\ - \sum_{j=0}^{T''''-t} \frac{r^j \left(c''''_{j+t} \times n''_{j+t} + c'''''_{j+t} \right)}{(1+i)^{j+t}} \end{array} \right] \right\}$
Berger and Nasr (1998)	*Assumptions:* ▪ Sales take place once a year; ▪ Yearly customer revenues, customer retention costs and the customer retention rate are constant over time; Furthermore, when the yearly net contribution margin per customer is constant over time, they also assume that: ▪ Revenues and the cost of sales (*i.e.*, the gross contribution margin) take place at the time of sale; ▪ The first transaction occurs at the time of the CLV computation (moment of customer acquisition); ▪ All promotional expenses occur at the middle of the purchase cycle. *(i) Considers all assumptions* $CLV = \left\{ GC \times \sum_{t=0}^{T} \frac{r^t}{(1+i)^t} \right\} - \left\{ c \times \sum_{t=1}^{T} \frac{r^{t-1}}{(1+i)^{t-0.5}} \right\}$ *(ii) Time period are shorter than one year, but equal in length* $CLV = \left\{ GC_s \times \sum_{t=0}^{n'T} \frac{(r_s)^t}{(1+i)^{t/n'}} \right\} - \left\{ c_s \times \sum_{t=1}^{n'T} \frac{(r_s)^{t-1}}{(1+i)^{(t-0.5)/n'}} \right\}$ *(iii) Time period are longer than one year, but equal in length* $CLV = \left\{ GC_s \times \sum_{t=0}^{T/n''} \frac{(r_s)^t}{(1+i)^{t\,n''}} \right\} - \left\{ c_s \times \sum_{t=1}^{T} \frac{(r_s)^{(t-1)/n''}}{(1+i)^{t-0.5}} \right\}$ *(iv) GC and M per customer are nonconstant over time* $CLV = \left\{ \sum_{u=0}^{T} \pi(u) \times \frac{r^u}{(1+i)^u} \right\}$

[2] The notations are presented in Appendice A.

Authors	CLV formula [2]
	(v) GC and M per customer are nonconstant over time and cash flows are continuous $$CLV = \pi(0) + \int_0^T \pi(u) \times \left(\frac{r}{1+i}\right)^u d(u)$$ *(vi) Noncontractual setting - repeat purchase behaviour* $$CLV = GC \times \left\{ n_b + \frac{1}{(1+i)^t} \times \sum_{t=1}^{T} \left[\sum_{j=1}^{t} n_{t-j} \times P_{u-j} \times \prod_{k=1}^{j} \left(1 - P_{u-j+k}\right) \right] \right\} \times \frac{1}{n_b}$$ $$- \frac{c}{(1+i)^{t+0.5}} \times \left\{ \frac{n_b}{(1+i)^{0.5}} + \left[\sum_{t=1}^{T} \sum_{j=1}^{t} n_{t-j} \times P_{u-j} \times \prod_{k=1}^{j} \left(1 - P_{u-j+k}\right) \right] \right\} \times \frac{1}{n_b}$$ with $P_u = 0$
Mulhern (1999)	$$CLV = \sum_{t=1}^{T'''} \frac{\sum_{j=1}^{n'} \left(R_{jt} - C_{jt}\right) - \sum_{k=1}^{n''''} c_{kt}}{(1+i)^t}$$
Hoekstra and Huizingh (1999)	$$CLV = \sum_{t=0}^{p} CQ_t \, (1+i)^{p-t} + \sum_{t=p+1}^{T'''} \left(CS_t \times CP_t\right)(1+i)^{p-t}$$ where, $t = 0$ - start date of the relationship with customer $t = p$ - current time period $t = T'''$ - projected ending of the relationship
Mani *et al.* (1999)	$$CLV = \sum_{t=1}^{T} S(t) \times v(t)$$
Reinartz and Kumar (2000)	$$CLV = \sum_{t=1}^{T} \frac{GC_t - c_t''}{(1+i)^t}$$
Pfeifer and Carraway (2000)	*Assumptions:* ▪ Transactions take place at most once a period and at the end of the period ▪ The probability of a transaction is a function only of the customer recency, *i.e.*, the number of periods since the last transaction. *(i) Finite horizon* $$CLV^T = \sum_{t=0}^{T} \left[(1+i)^{-1} S \right]^t CF^V$$ *(ii) Infinite horizon* $$CLV^T = \left\{ I - (1+i)^{-1} S \right\}^{-1} CF^V$$
Hogan *et al.* (2002a)	*Assumptions:* ▪ Defection rate is constant over time and across customers; ▪ Customer margin (profits minus retention costs) is constant over time; ▪ Growth rate is constant over time.

Authors	CLV formula [2]
	(i) Current customers continue to buy the same thing from the firm $$CLV = \frac{m}{i+d}$$ *(ii) Up-selling and cross-selling strategies are adopted* $$CLV = \frac{m}{i+d-g} \qquad , g < i+d$$
Calciu and Salerno (2002)	*Assumptions:* ▪ Profits per transaction are constant over time; ▪ Gains and sales expenses occur at the same time. <u>*Retention model*</u> *(i) Retention probability varies over time* $$CLV = (R'-C')\sum_{k=0}^{T'''}\left[\prod_{t=0}^{k}\frac{r_t}{(1+i)^k}\right]$$ *(ii) Retention probability is constant over time* $$CLV = (R'-C')\sum_{k=0}^{T'''}\left[\frac{r}{(1+i)}\right]^k$$ *(iii) Retention probability is constant over time and the horizon is infinite* $$CLV = (R'-C')\times\frac{(1+i)}{(1+i-r)}$$ *(iv) Retention probability is a function of the marketing effort (budget) directed towards the customer during each period and the horizon is infinite* $$CLV = \left(R'-\frac{c'}{r_c}\right)\times\frac{(1+i)}{(1+i-r_c)}$$ <u>*Migration model*</u> $$CLV_{r'} = \sum_{k=0}^{T'''}\frac{m'''\times n_{r',k} - c''\ (1-Q_k)}{(1+i)^k} \qquad , where\ Q_k = \sum_{l=0}^{k} q_k$$ and $$q_k = \begin{cases} n_{1,k-r''} \prod_{l=1,k\geq r''}^{r''}(1-p_l),\ r'=1 \\ \prod_{l=1,\ k=r''-r'+1}^{r''}(1-p_l)+n_{r',k-r''}\prod_{l=1,\ k\geq r''+1}^{r''}(1-p_l),\ r'>1 \end{cases}$$
Gurau and Ranchhod (2002)	*Assumptions:* ▪ Margin is constant over time and across customers; ▪ Lifetime is constant across customers; ▪ Acquisition cost is constant across customers; ▪ Number of customers is constant over time.

Authors	CLV formula [2]
	$CLV = (R - C) \times T''' - A$
Rosset *et al.* (2002)	$CLV = \int_0^{\infty} S(t) v'(t) D(t)\, dt$
Bayón *et al.* (2002)	$CLV = \sum_{t=0}^{T'''} \frac{m_{dt} + m_{WoM}}{(1+i)^t} \times W \qquad , W \geq 1$
Bauer *et al.* (2003)	$CLV = -A + \sum_{t}^{T} \left((r_t)^t \times \frac{\left(R_t^A + R_t^{US} + R_t^{CS} + GC'_t\right) - \left(C_t''' + c_t\right)}{(1+i)^t} - \left[(r_t)^{t-1} \times (1 - r_t)\right] \times \frac{c_t'''}{(1+i)^t} + (r_t)^t \times \left\{ \frac{InfoV_t + CoopV_t + InnoV_t}{(1+i)^t} \right\} \right)$
Reinartz and Kumar (2003)	$CLV = \sum_{t=1}^{T} S(t) \times CM_t \left(\frac{1}{1+i} \right)^t$
Gupta and Lehmann (2003)	*(i) Retention probability varies over time and the time horizon is finite* $CLV = \sum_{t=1}^{T'''} \frac{m_t \times \prod_{j=1}^{t} r_j}{(1+i)^t}$ *(ii) Customer margin and retention probability are constant over time and the time horizon is infinite* $CLV = m' \left(\frac{r}{1+i-r} \right)$ *(iii) Margin increases at a constant rate (g)* $CLV = m' \left(\frac{r}{1+i-r\,(1+g)} \right)$ *(iv) Margin increases at a decreasing rate* $CLV = m_0 \left(\frac{r}{1+i-r} \right) \left(\frac{(1+i-r) + s\,r\,\frac{m_\infty}{m_0}}{1+i-(1-s)r} \right)$ *(v) Retention rate increases over time* $r_t = r_0 + (r_\infty - r_0)\,[1 - \exp(-s't)]$ *(vi) Customer lifetime is finite (n)* $CLV = m' \left(\frac{r}{1+i-r} \right) \left[1 - \left(\frac{r}{1+i} \right)^T \right]$

Authors	CLV formula [2]
Gupta *et al.* (2004)	*Assumptions:* ▪ Process is continuous; ▪ Retention rate is constant. $CLV = \sum_{t=0}^{\infty} m_t \frac{r^t}{(1+i)^t}$
Malthouse and Blattberg (2004)	$CLV = \sum_{t=1}^{T} \frac{m_t}{(1+i)^t}$
Pfeifer and Farris (2004)	$CLV = \sum_{t=1}^{\infty} \left(\frac{1}{1+i}\right)^t \left(\prod_{j=1}^{t} r_j\right)(m_t - c'_t)$
Hwang *et al.* (2004)	$CLV = \underbrace{\sum_{t=0}^{T'} \frac{\pi_p(t)}{(1+i)^{t-T'}}}_{\text{Past profit contribution}} + \underbrace{\sum_{t=T'+1}^{T'+T''+1} \frac{\pi_f(t)+B(t)}{(1+i)^{t-T'}}}_{\text{Expected future cash flow}}$
Kumar *et al.* (2004)	*Assumptions:* ▪ The average acquisition cost per customer, the average gross contribution and marketing costs per customer are constant over time; ▪ The number of acquired customers per period k is constant; ▪ The retention rate is constant over time; ▪ The cost of capital is fixed. *(i) Average CLV - considers all assumptions* $CLV = \sum_{k=0}^{\infty}\left[\frac{1}{(1+i)^k}\left\{\sum_{t=k}^{\infty}\left[\frac{(GC'-c''')}{(1+i)^{t-k}} r^{t-k}\right] - A'\right\}\right]$ *(ii) Average CLV - the number of acquired customers per period is not constant* $CLV = \frac{1}{\sum_{k=0}^{\infty} n_k}\left[\sum_{k=0}^{\infty}\frac{n_k}{(1+i)^k}\sum_{t=k}^{\infty}\left\{\frac{(GC'-c''')}{(1+i)^{t-k}} r^{t-k} - A'\right\}\right]$ *(iii) Average CLV - the number of acquired customers per period is not constant; G and M'' are not constant over time and for every cohort; and A are not constant over time* $CLV = \frac{1}{\sum_{k=0}^{\infty} n_k}\left[\sum_{k=0}^{\infty}\frac{n_k}{(1+i)^k}\sum_{t=k}^{\infty}\left\{\frac{(GC'_{t-k}-c'''_{t-k})}{(1+i)^{t-k}} r^{t-k}\right\} - \sum_{k=0}^{\infty}\frac{n_k\ A'_k}{(1+i)^k}\right]$
Venkatesan and Kumar (2004)	*Assumptions:* ▪ The amount of customer spending is independent of purchase timing. $CLV = \sum_{y=1}^{n'''} \frac{CM_y}{(1+i)^{y/f}} - \sum_{t=1}^{T} \frac{\sum^{m} c_{m,t} \times n_{m,t}}{(1+i)^{t-1}}$

Authors	CLV formula [2]
Pfeifer *et al.* (2005)	*(i) Expected lifetime value of a prospect* $CLV = -A + a'm + a'\left(m - \frac{c'}{r}\right)\left[\frac{D\,r}{1 - D\,r}\right]$ *(ii) Expected lifetime value of a new customer* $CLV = m + \left(m - \frac{c'}{r}\right)\left[\frac{D\,r}{1 - D\,r}\right]$ *(iii) Expected lifetime value of a just-acquired customer* $CLV = \left(m - \frac{c'}{r}\right)\left[\frac{D\,r}{1 - D\,r}\right]$
Ryals (2005)	$CLV = \sum_{t=1}^{T'''} \frac{R_t - C_t}{(1+i)^t}$
Yang (2005)	$CLV\ on\ average = \sum_{t=1}^{T} \frac{(R_t' - C_t')}{(1+i)^t \times n_0}$
Bauer and Hammerschmidt (2005)	*Assumptions:* ▪ Sales occur annually; ▪ Revenues and costs may vary over time, but within a year all cash flows are discrete and take place at the end of each purchase cycle; ▪ Retention rate is constant over time. $CLV = -A + \sum_{t}^{T}\left(r^t \frac{\left(R_t^A + R_t^{US} + R_t^{CS} + GC'_t\right) - \left(C_t''' + c_t\right)}{(1+i)^t} - \left(r^{t-1}(1-r)\right)\frac{c_t'''}{(1+i)^t}\right)$
von Wangenheim (2006)	$CLV = \sum_{t=1}^{T} \frac{\left(CM_{NRT} \times n_t + CM_{Up} \times n'_t\right)}{(1+i)^t}$
Fader and Hardie (2006)	$CLV = \sum_{t=0}^{\infty} m \frac{S(t)}{(1+i)^t}$
Gupta and Zeithaml (2006)	*(i) Considers the expected lifetime estimated based on a retention model* $CLV = \sum_{t=0}^{T'''} \frac{\left(R_t - C_t''\right)}{(1+i)^t} - A$ *(ii) Including the probability of retention in the equation* $CLV = \sum_{t=0}^{T} \frac{\left(R_t - C_t''\right) r_t}{(1+i)^t} - A$

From an analysis of the CLV formulas proposed in the literature, it can be concluded that the most common components are: (i) cash flow, (ii) retention rate, (iii) time horizon, and (iv) discount rate. Some researchers argue that one of the most important components of customer value is the retention probability of the customer at each period, which should influence the customer cash flows. The retention probability is the probability that the customer continues to do business with the same service/product provider.

The cash flow concept has not been accurately applied to CLV, as many 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 propose the contribution margin[3] (*e.g.,* Berger and Nasr, 1998; Malthouse and Blattberg, 2004; Reinartz and Kumar, 2000) or the margin (*e.g.,* Gupta *et al.*, 2003, 2004; Hogan *et al.*, 2002).

The majority of researchers use a finite horizon in the CLV prediction. But, while some researchers argue that time horizon should equals the customer lifetime (*e.g.,* Gupta and Zeithaml, 2005; Hwang *et al.*, 2004; Jain and Singh, 2002; Villanueva, 2003; Wyner, 1996), others use an arbitrary time horizon (which is usually a short period, between 3 to 5 years) (*e.g.,* Donkers *et al.*, 2003; Pfeifer and Bang, 2005; Reinartz and Kumar, 2000, 2003; Rust *et al.*, 2004).

It should be noted that it is usual to find proposals of CLV formulas based on assumptions that are misadjusted to the business reality as well as to the financial theory of assets evaluation. One of the most used assumptions is the constant contribution margin over time (*e.g.,* Bonfrer *et al.*,2007; Calciu and Salerno, 2002; Donkers *et al.*, 2003; Drew *et al.*, 2001; Hogan *et al.*, 2002a; Kumar *et al.*, 2004) and across customers (*e.g.,* Gupta and Lehmann, 2003; Gurau and Ranchhold, 2002).

In addition, many researchers also assume a constant retention rate over time (*e.g.,* Bauer and Hammerschmidt, 2005; Berger and Nasr, 1998; Blattberg and Deighton, 1996; Gupta and Lehmann, 2003; Gupta *et al.*, 2004; Hogan *et al.*, 2002a; Kumar *et al.*, 2004) and

[3] According to the accounting theory, the contribution margin is the difference between revenues and variable costs.

across customers (Hogan *et al.*, 2002a). These assumptions will be empirically analysed in sections 4.3 e 4.4 for the fixed-telephone and ADSL contracts.

CLV modulation has been widely criticised in the literature, mainly due to the incapacity of encompassing all variables that affect customer behaviour (*e.g.*, Bauer *et al.*, 2003; Bauer and Hammerschmidt, 2005; Lee and Park, 2005). The majority of the CLV models proposed in the literature are deterministic. However, Villanueva and Hanssens (2007) emphasise the numerous advantages of using stochastic models of CLV and suggest that more research is needed on this direction.

2.6. Customer churn

2.6.1. Customer churn definition

The differences between contractual and noncontractual relationships lead to different concerns on the customer management, especially on the customer retention management. For example, in contractual settings, customers must contact service/product providers in order to cancel the contract; and, thus, firms know when each customer defects. So, the firm knows exactly who are the active and inactive customers. Hence, firms' uncertainty is about the probability of (active) customer defection over time. On the other hand, in noncontractual settings, firms do not know which customers are active or not, because they do not contact the firms to terminate the relationship. So, in this case, firms' main concerns are twofold: (i) which customers are still active at the moment; and (ii) within the active customers, what are the predictions of future transactions. According to the reasons mentioned above, the retention probability only makes sense in contractual settings (Schmittlein *et al.*, 1987; Schmittlein and Peterson, 1994). As the present study intends to analyse the relationship with customers with contracts of fixed-telephone or ADSL with the service provider, this section is only about customers' relationships in contractual settings. Nevertheless, it should be noted that a customer that terminates a given contract with the firm is not considered lost forever.

The complement of customer retention is customer defection, commonly designated by customer churn in the telecommunications industry. Customer churn reflects the customer's decision to terminate the relationship with a provider, either because the customer does not need its products or services anymore or because the customer wants to switch to another product/service provider.

There are different types of customer churn, depending on the agent who cancels the relationship. Thus, some authors argue that customer churn may be voluntary or involuntary (Desai, 2006; Hadden *et al.*, 2005; Li *et al.*, 2006; Lu, 2002). Berry and Linoff (2004) classify the customer churn into three categories: voluntary churn, involuntary churn and expected churn. Customer churn is voluntary when the relationship is cancelled by decision of the customer. On the other hand, customer churn is involuntary when the provider decides to terminate the relationship with the customer (usually due to missed payments, bad debts, etc.). Berry and Linoff (2004) define expected churn as the end of the relationship due to the fact that the customer is no more on the target market for a product/service. According to Burez and Van den Poel (2008), there are four types of customer churn: (i) involuntary churn (customers who died or moved abroad); (ii) financial churn (customers who stop paying the service due to financial concerns); (iii) commercial churn (customers who cancel the service because they do not want it anymore); and overall churn (customers who churn due to a mix of financial and commercial reasons). Researchers have been focused on voluntary customer churn.

Pettersson (2004) proposes another classification of churn. According to him, churn can be total (a customer who completely stops buying from the firm) or partial (a customer who cancel at least one product/service, but still buys other products/services from the firm).

2.6.2. Studies about customer churn prediction

The customer churn issue is present both on studies about CLV (as a component of this metric) and on specific studies about churn, using different perspectives. In studies about CLV, customer churn is mainly analysed in a theoretical way and its prediction is usually neglected, whereas on the later case, the statistical models with empirical data are predominant. Most studies that focus on CLV make strong assumptions about customer retention, such as customer retention is constant over time (*e.g.*, Berger and Nasr, 1998; Blattberg and Deighton, 1996; Gupta and Lehmann, 2003; Gupta *et al.*, 2004; Hogan *et al.*, 2002; Kumar *et al.*, 2004) and/or across customers (Hogan *et al.*, 2002). Schweidel *et al.* (2006) point out that these assumptions are unrealistic because they omit the different customer switching behaviour over time, the duration dependence, the influences of marketing activities on customers and the heterogeneity across customers' churn propensities. Nevertheless, the limitations of these assumptions are not recognised by all researchers. Although these assumptions simplify the calculations and allow for the non-development of churn prediction models, they do not mirror the customer behaviour. In their study about prediction of customer retention in telecommunications, Schweidel *et al.* (2006) found that constant retention rates over time lead to lower expected customer tenure than the tenure they predict with survival analysis. They also found that neglecting customer heterogeneity can induce to large errors in churn prediction.

Customer churn prediction is a more recent area of research than the study of CLV. The majority of studies about customer retention focus on specific determinants of customer retention (for instance, customer satisfaction), and on the consequences of high retention rates (for instance, on firm profitability). Studies about customer churn prediction have only flourished in the last years, mainly in the telecommunications industry, due to the high churn rates that have characterised this industry.

Villanueva and Hanssens (2007) mention that, according to the published literature, the three most important determinants of customer retention are: (i) switching costs, (ii) customer satisfaction, and (iii) customer future considerations about usage.

Many researchers have focused on the analysis of the effect of customer satisfaction on customer retention and the most of them found evidence that customer satisfaction positively influences customer retention. For instance, Bolton (1998) found a positive effect of overall customer satisfaction on the duration of the relationship with the customer, in mobile telecommunications industry. Ranaweera and Prabhu (2003) found evidence that customer satisfaction has a strong positive association with customer retention in the fixed telephone industry in UK. Gustafsson *et al.* (2005) also found a positive effect of customer satisfaction on customer retention, in telecommunications industry. Nevertheless, Jones and Sasser (1995) and Chandrashekaran *et al.* (2007) point out that customer satisfaction alone could not be enough to ensure the customer retention, because there are satisfied customers that defect.

Customer churn has been studied using different techniques, in different industries (*e.g.*, banking, insurance, telecommunications), and in different contexts (contractual vs. noncontractual settings, continuous vs. discrete time). Buckinx and Van den Poel (2005), Hadden *et al.* (2005), Mutanen (2006), Song *et al.* (2004), and Van den Poel and Larivière (2004) present reviews of literature about customer churn. Ahn *et al.* (2006) point out that the reasons of customer churn and the customer behaviour towards churn need to be more studied. The following table presents a review of the literature about customer churn prediction in the telecommunications industry.

Table 3 – Summary of the literature review on customer churn prediction in the telecommunications industry

Authors	Scope of the study	Industry	Country / Region	No. of observations	Timing range of the data	Data collection	Statistical technique	Number of covariates
Ahn *et al.* (2006)	- Development of a comprehensive customer churn model (residential customers) - Analysis of the mediating effects of a customer's partial defection on the relationship between the churn determinants and total defection	Mobile telecommunications	South Korea	5 789 (652 churn)	8 months	Internal database	Binary logistic regression	15
Bin *et al.* (2007)	Customer churn prediction	Personal handyphone system service (PHSS)	China	4 799 (737 churn)	180 days	Internal database	Decision trees	3
Bonfrer *et al.* (2007)	Examine the degradation process (usage rate over time), at individual level and before the defection event occurs	Mobile telecommunications	China	1 662 (114 churn)	12 months	Internal database	- Arithmetic Brownian motion (ABM) - Geometric Brownian motion (GBM)	n.a.
Burez and Van den Poel (2007)	- Customer churn prediction by using different statistical techniques - Customer targeting - Analysis of three different customer retention strategies	Pay-TV	Europe	n.a.	n.a.	Internal database	- Binary logistic regression - Markov chains - Random forests	31
Burez and Van den Poel (2008)	Customer churn prediction at a specific moment in time (static) and over time (dynamic)	Pay-TV	Europe	Over 500 000 (dynamic) 100 000 (static)	n.a.	Internal database	- Survival analysis (dynamic) - Random forests (static)	0(dynamic) 171 (static)
Chen *et al.* (2007)	- Customer segmentation based on customer trend - Survival analysis of each cluster	Telecommunications	China	1 000	n.a.	Internal database	- K-means clustering arithmetic (data mining) - Survival analysis (data mining)	196
Drew *et al.* (2001)	Estimate the customer's hazard function	Telecommunications	US	21 500	n.a.	Internal database	- Classical survival analysis - Artificial neural networks for survival analysis	n.a.
Eshghi *et al.* (2007)	Investigate the propensity to switch the service provider	Mobile telecommunications	US	2 861	n.a.	Phone survey	Structural equation model	30

Authors	Scope of the study	Industry	Country / Region	No. of observations	Timing range of the data	Data collection	Statistical technique	Number of covariates
Ferreira *et al.* (2004)	- Customer churn prediction - Comparison of the predictive and explanatory power of four types of models	Mobile telecommunications	Brazil	100 000 (1 245 churn)	9 months	Internal database	- Multilayer perceptron neural networks - C4.5 decision trees - Hierarchical neuro-fuzzy systems - Rule evolver (based on genetic algorithms)	37
Hung *et al.* (2006)	Compare several data mining techniques that can assign a "propensity-to-churn" score periodically for each customer	Mobile telecommunications (post-paid subscribers)	Taiwan	160.000 (14 000 churn)	12 months	Internal database	- K-means clustering - Decision tree - Back propagation neural networks	10
Jamal and Bucklin (2006)	- Customer churn prediction - Study the link between customer churn and some characteristics of the customer behaviour	Direct-to-home satellite TV	South America	2 801	12 months	Internal database	Latent class Weibull hazard model	10
Kim and Yoon (2004)	Identify the determinants of customer churn and customer loyalty	Mobile telecommunications	Korea	973	n.a.	Phone survey	Binary logistic regression	14
Lemmens and Croux (2006)	Analyse if the bagging and boosting classification techniques outperform the binary logistic model in predicting churn	Mobile telecommunications	US	203 074	n.a.	Teradata Center at Duke University	- Binary logistic regression - Bagging - Stochastic gradient boosting	171 (46 after a reduction process)
Li *et al.* (2006)	Customer churn prediction	Telecommunications	China	40 000	6 months	Internal database	Data mining	110 (61 after a reduction process)
Lu (2002)	Customer churn prediction	Telecommunications	n.a.	41 374	15 months	Internal database	Survival analysis (data mining)	212 (29 after a reduction process)
Lu (2003)	- Development of the concept of CLV - Demonstrate how survival analysis techniques are used in the estimation of CLV	Telecommunications	n.a.	64 320	20 months	Internal database	Survival analysis (data mining)	42

Authors	Scope of the study	Industry	Country / Region	No. of observations	Timing range of the data	Data collection	Statistical technique	Number of covariates
Madden *et al.* (1999)	Customer churn prediction (residential users)	ISP	Australi a	592	n.a.	Web based survey	Binomial probit	19
Mani *et al.* (1999)	Modelling customer lifetime	Mobile telecommunications	US	21 500	n.a.	Internal database	- Classical survival analysis - Neural networks for survival analysis	40
Mozer *et al.* (2000a)	- Customer churn prediction (residential customers) - Identify customers to whom incentives should be offered to increase retention	Mobile telecommunications	US	46 744	3 months	Internal database	- Binary logistic regression - Neural networks	134
Mozer *et al.* (2000b)	- Customer churn prediction (residential customers) - Determine what incentives should be offered to customers in order to improve the retention and maximize the profitability of the firm	Mobile telecommunications	US	46 744	3 months	Internal database	- Binary logistic regression - Decision trees - Neural networks - Boosting	134
Nath and Behara (2003)	Customer churn prediction	Mobile telecommunications	n.a.	About 100 000	n.a.	Teradata Center at Duke University	Naïve Bayes algorithm data mining option for supervised learning	171
Neslin *et al.* (2006)	Identify the best approach in the prediction of customer churn	Mobile telecommunications	n.a.	About 100 000	n.a.	Teradata Center at Duke University	- Binary logistic regression - Decision trees - Neural networks - Discriminant analysis	171
Qian *et al.* (2006)	Profile customer behaviour in order to identify and capture churn activity patterns	Telecommunications	n.a.	1 787	24 months	Internal database	Functional mixture model	24
Rosset and Neumann (2003)	Customer churn prediction	Telecommunications	n.a.	3 000 (1 500 churn)	n.a.	Internal database	Binary logistic regression (data mining)	400
Rosset *et al.* (2002)	Calculation and discussion of the potential applications of CLV and its components	Mobile telecommunications	n.a.	n.a.	n.a.	n.a.	Churn Management System from Amdocs' BI platform (Automatic knowledge discovery)	n.a.

Authors	Scope of the study	Industry	Country / Region	No. of observations	Timing range of the data	Data collection	Statistical technique	Number of covariates
Schweidel *et al.* (2006)	Modelling retention within and across cohorts in contractual relationships	Telecommunications	n.a.	n.a.	29 months	Internal database	Survival analysis	3
Seo *et al.* (2007)	Examine the influence of some variables on the customer retention behaviour	Mobile telecommunications	US	30 572 (14 068 churn)	6 months	Internal database	- Binary logistic regression - Hierarquical linear model	6
Wei and Chiu (2002)	Customer churn prediction	Mobile telecommunications	Taiwan	114 000 (4 500 churn)	4 months	- Interviews - Internal database	Decision trees	9
Yan *et al.* (2001)	Propose two distinct approaches o improve the performance of churn prediction models in nonstationary environments (business customers)	Mobile telecommunications	US	About 70 000	6 months	Internal database	Multilayer perceptron neural networks	71
Zhang *et al.* (2006)	Customer churn prediction	Fixed-line telephone	China	17 223 (5 167 churn)	7 months	Internal database	Different data mining technologies (decision trees, neural networks, and regression)	17

From the analysis of Table 3, it can be concluded that despite the extensive research done on customer churn, there are few studies applied to the ISP and fixed-telephone industries. On the contrary, the majority of published research about customer churn prediction in the telecommunications industry analyses the mobile telecommunications. The most studied geographic areas are Asia and the U.S.. This issue has never been studied in Portugal. Many studies focus on model accuracy or comparison of techniques rather than on testing the effect of churn covariates. The data were mostly gathered from internal databases. Most of these studies were presented in data mining conferences; so, the majority of them apply data mining techniques to predict customer churn. Lastly, most of them model if a customer is likely to churn at a pre-specified time period, rather than the longitudinal churn pattern over the duration of the relationship. Voss (1998) also mentions that there are few studies modelling survival time in the marketing literature.

Apart from these studies, there are others that focus on finding the relationship between a variable or a little set of variables and customer retention, *i.e.*, the determinants of customer retention (*e.g.*, Bolton, 1998; Capraro *et al.*, 2003; Gerpott *et al.*, 2001; Yu *et al.*, 2005).

In this context, it seems relevant to propose a more detailed study of the customer churn in the fixed telecommunications industry, as it may be misleading to make decisions based only on the results of the mobile telecommunications industry, which presents very different characteristics. Moreover, considering that the customer churn behaviour may be influenced by the customer culture, it is pertinent to examine different markets, like in our case, the Portuguese one. In this context, the aim of this study is to develop a model of the residential partial customer churn in the fixed-telecommunications industry in Portugal. It also intends to analyse the assumptions of constant retention rate over time and across customers.

2.6.3. Variables used in the estimation of customer churn

Table 4 presents a list of some variables used in the estimation of the probability of customer churn. The variables are grouped into different categories, namely: contract, usage, revenues, promotions, switching costs, payment history, equipment, technology intensity, satisfaction, CRM, market, socio-demographic, economic, ownership, and other. The most used variables are the duration of the contract, the rate plan, the monthly revenues, customer age, gender and the geographic area. It should be noted that many researchers do not mention the covariates used in their studies or they only indicate some of them. As can be seen in Table 4, few researchers indicate which covariates are significant and the sign of their effect on customer churn. Burez and Van den Poel (2008) present a list of all variables used in their customer churn prediction model, which is an exception. Most researchers do not incorporate covariates about business competition on their models because of the inexistence of available data about this issue (Bell *et al.*, 2002; Gupta *et al.*, 2006).

Table 4 – Summary of some variables used in the customer churn estimation in the telecommunications industry

		Ahn *et al.* (2006)	Bin *et al.* (2007)	Bonfrer *et al.* (2007) [1]	Burez and Van den Poel (2007)	Chen *et al.* (2007) [2]	Drew *et al.* (2001)	Eshghi *et al.* (2007)	Ferreira *et al.* (2004) [2]	Hung *et al.* (2006) [3]	Jamal and Bucklin (2006)	Kim and Yoon (2004)	Lemmens and Croux (2006) [2]	Li *et al.* (2006) [2]	Lu (2002) [2]	Lu (2003) [1]	Madden *et al.* (1999)	Mani *et al.* (1999) [2]	Mozer *et al.* (2000a) [2]	Mozer *et al.* (2000b) [2]	Nath and Behara (2003) [2]	Neslin *et al.* (2006) [2]	Qian *et al.* (2006)	Rosset and Neumann (2003) [1]	Rosset *et al.* (2002) [1]	Schweidel *et al.* (2006b)	Seo *et al.* (2007)	Wei and Chiu (2002)	Yan *et al.* (2003) [1]	Zhang *et al.* (2006)
Contract	Duration of the contract				x	x	x			x*		x-	x					x				x				x	x-	x		x
	Dates of each contract (begin / finish)																		x	x	x									
	Customer status	x+																												
	Month of contract expiration				x																									
	Contract type						x											x										x		
	Payment method	x			x						x																	x		x
	Rate plan	x					x		x					x	x		x	x	x	x	x									
	Service plan complexity																										x-			
	Toll free services													x	x															
	Change account (phone number, etc.)									x*																				
	Active products or services (by type)				x				x		x				x															
	Number of active products/ services																		x	x	x									
Usage	Call details									x*									x	x	x									
	Monthly usage													x					x	x	x									
	Total number of outbound calls		x																											
	Total number of different numbers contacted		x																											
	Duration of outbound calls (in minutes)		x																											
	Weekly average call accounts													x	x															
	Percentage change of minutes													x	x															
	Cumulative invoice amount										x																			
	Minutes of use in sub-periods																											x		
	Frequency of use in sub-periods																											x		
	Sphere of influence in sub-periods																											x		
	Variation of duration of calls (in minutes)																											x		
	Variation of number of calls																											x		
	Variation of different numbers contacted																											x		
	Duration of calls									x*																				
	Mean number of attempted calls												x									x								
	Variation in monthly duration of calls vs previous 3-months average												x									x								

		Ahn *et al.* (2006)	Bin *et al.* (2007)	Bonfrer *et al.* (2007) [1]	Burez and Van den Poel (2007)	Chen *et al.* (2007) [2]	Drew *et al.* (2001)	Eshghi *et al.* (2007)	Ferreira *et al.* (2004) [2]	Hung *et al.* (2006) [3]	Jamal and Bucklin (2006)	Kim and Yoon (2004)	Lemmens and Croux (2006) [2]	Li *et al.* (2006) [2]	Lu (2002) [2]	Lu (2003) [1]	Madden *et al.* (1999)	Mani *et al.* (1999) [2]	Mozer *et al.* (2000a) [2]	Mozer *et al.* (2000b) [2]	Nath and Behara (2003) [2]	Neslin *et al.* (2006) [2]	Qian *et al.* (2006)	Rosset and Neumann (2003) [1]	Rosset *et al.* (2002) [1]	Schweidel *et al.* (2006b)	Seo *et al.* (2007)	Wei and Chiu (2002)	Yan *et al.* (2003) [1]	Zhang *et al.* (2006)
	Average monthly duration of calls over the previous 6 months												x									x								
	Mean number of completed calls												x									x								
	Mean number of peak calls												x									x								
	Mean number of inbound calls < 1 minute												x									x								
	Monthly mean duration of calls												x									x								
	Mean total duration of outbound wireless to wireless calls (in minutes)												x									x								
	Total number of calls						x											x												
	Duration of local, peak, and off-peak calls						x											x												
Revenues	Monthly revenues	x+			x				x	x*		x	x				x+		x	x	x	x	x							
	Mean revenues over the customer lifetime												x									x								
	Monthly fee									x*																				
	Roaming								x																					
	Value of minutes used						x											x												
	Value of toll						x											x												
	Value of roaming and optional features						x											x												
	Current and historical profitability						x											x												
	Proportion of service fee i																													x
	Consumption level rate[4]																													x
	Growth rate of total fee																													x
	Growth rate of consumption level rate																													x
	Quantity abnormal fluctuation of total fee																													x
	Quantity of abnormal fluctuation of consumption level rate																													x
	Share of domestic / international revenues													x	x															
	Average revenues												x									x								
Promotions	Total discount in the past 6 months					x																								
	Discount													x	x															

[4] Revenues of customer i in month j / revenues of all customers in month j

		Ahn *et al.* (2006)	Bin *et al.* (2007)	Bonfrer *et al.* (2007) [1]	Burez and Van den Poel (2007)	Chen *et al.* (2007) [2]	Drew *et al.* (2001)	Eshghi *et al.* (2007)	Ferreira *et al.* (2004) [2]	Hung *et al.* (2006) [3]	Jamal and Bucklin (2006)	Kim and Yoon (2004)	Lemmens and Croux (2006) [2]	Li *et al.* (2006) [2]	Lu (2002) [2]	Lu (2003) [1]	Madden *et al.* (1999)	Mani *et al.* (1999) [2]	Mozer *et al.* (2000a) [2]	Mozer *et al.* (2000b) [2]	Nath and Behara (2003) [2]	Neslin *et al.* (2006) [2]	Qian *et al.* (2006)	Rosset and Neumann (2003) [1]	Rosset *et al.* (2002) [1]	Schweidel *et al.* (2006b)	Seo *et al.* (2007)	Wei and Chiu (2002)	Yan *et al.* (2003) [1]	Zhang *et al.* (2006)
	Promotion													x	x											x				
Switching costs	Membership card	x+																												
	Loyalty points	x -																												
	Total offers accepted from retention team																				x*									
Payment history	Credit classification																		x	x	x									
	Count of bar or suspended									x*																				
	Number of overdue accounts				x					x*																				
	Value of overdue accounts	x-			x						x																			
	Number of reminders				x																									
	Type of reminders				x																									
	Elapsed time since the last reminder				x																									
	Number of unpaid monthly bills in time	x																												
Equipment	Age								x			x+	x									x								
	Capability (internet)	x -							x																					
	Manufacturer	x -																												
	Type				x														x	x	x*									
	Sophistication																										x -			
	Price												x									x								
Technology intensity	Time on the internet per week							x																						
	Years connected to the internet							x																						
	Internet use variables																x													
	ISP choice variables																x													
	Level of high technology involvement							x																						
	Satisfaction with call quality											x -																		
	Satisfaction with tariff level											x-																		
Satisfaction	Satisfaction with billing											x																		
	Satisfaction with value-added services											x																		
	Satisfaction with customer services											x																		
	Satisfaction with handset											x-																		
	Satisfaction with brand image											x-																		

		Ahn *et al.* (2006)	Bin *et al.* (2007)	Bonfrer *et al.* (2007) (1)	Burez and Van den Poel (2007)	Chen *et al.* (2007) (2)	Drew *et al.* (2001)	Eshghi *et al.* (2007)	Ferreira *et al.* (2004) (2)	Hung *et al.* (2006) (3)	Jamal and Bucklin (2006)	Kim and Yoon (2004)	Lemmens and Croux (2006) (2)	Li *et al.* (2006) (2)	Lu (2002) (2)	Lu (2003) (1)	Madden *et al.* (1999)	Mani *et al.* (1999) (2)	Mozer *et al.* (2000a) (2)	Mozer *et al.* (2000b) (2)	Nath and Behara (2003) (2)	Neslin *et al.* (2006) (2)	Qian *et al.* (2006)	Rosset and Neumann (2003) (1)	Rosset *et al.* (2002) (1)	Schweidel *et al.* (2006b)	Seo *et al.* (2007)	Wei and Chiu (2002)	Yan *et al.* (2003) (1)	Zhang *et al.* (2006)
	Global satisfaction with the firm										x																			
	Call drops rate	x^{+}																	x	x	x						x^{+}			
	Call failure rate	x																												
	Billing dispute														x															
	Quality of service data																		x	x	x									
	Problem solving																		x	x	x									
CRM	Customer segment	x													x															
	Contacts from the firm																													
	- Surveys				x																									
	- Letters by type				x																									
	- Calls by type				x																									
	- Mail													x	x															
	- Cumulative number of contacts										x																			
	- Respond to an offer sent by mail (y/n)												x									x								
	Contacts from the customer																													
	- Number of calls by type (general, requests to change service, inquiry about cancellation, complaints, etc)													x	x				x	x	x									
	- Mean duration of calls												x									x								
	- Number of complaints	x^{+}																												
Market	Competitor rates								x																					
	Advertising costs								x																					
	Business/ residential				x																									
	Age				x	x	x	x	x	x*	x	x		x			x^{-}	x									x^{-}			
	Gender	x^{+}			x	x		x	x	x*	x	x			x		x^{+}										x^{-}			
Socio-Demographic	Marital status					x					x		x		x							x								
	Level of education					x		x													x*									
	Ethnicity																				x*									
	Geographic area				x				x				x						x	x	x*	x								
	Years at current address					x																								
	Years with current employer					x																								

		Ahn *et al.* (2006)	Bin *et al.* (2007)	Bonfrer *et al.* (2007) (1)	Burez and Van den Poel (2007)	Chen *et al.* (2007) (2)	Drew *et al.* (2001)	Eshghi *et al.* (2007)	Ferreira *et al.* (2004) (2)	Hung *et al.* (2006) (3)	Jamal and Bucklin (2006)	Kim and Yoon (2004)	Lemmens and Croux (2006) (2)	Li *et al.* (2006) (2)	Lu (2002) (2)	Lu (2003) (1)	Madden *et al.* (1999)	Mani *et al.* (1999) (2)	Mozer *et al.* (2000a) (2)	Mozer *et al.* (2000b) (2)	Nath and Behara (2003) (2)	Neslin *et al.* (2006) (2)	Qian *et al.* (2006)	Rosset and Neumann (2003) (1)	Rosset *et al.* (2002) (1)	Schweidel *et al.* (2006b)	Seo *et al.* (2007)	Wei and Chiu (2002)	Yan *et al.* (2003) (1)	Zhang *et al.* (2006)
	Retired (y/n)					x																								
	Number of people in the household					x											x+													
	Children in the household (y/n)												x									x								
	Number of children														x															
	Children's age														x															
	Number of adults														x															
	Primary household member's age												x		x							x								
	Primary household member's occupation														x						x*									
	Account spending limit												x									x								
	Social group												x									x								
Economic	Income							x				x+	x				x+				x*	x								
	Household income and wealth ranking														x															
Ownership	Telephone (mobile, fixed phone, none)				x																									
	Car type																				x*									
	Type of residence										x																			
	Number of credit cards														x															
	Number of vehicles														x															
	Vehicles value														x															
Other	Previous switching experience (y/n)											x																		

(1) The predictors used in the study are not presented
(2) Only some predictors are presented
(3) Only significant predictors are presented

* Significant covariate
\+ Positive effect on customer churn
\- Negative effect on customer churn

2.6.4. Models to estimate customer churn

Several types of models have been used to estimate customer churn. Some of them are: heuristic models (for instance, the RFM model), the Schmittlein *et al.* (1987) model and its extensions, binary logistic models, discriminant analysis, survival models, decision trees, and artificial neural networks (ANN). It should be noted that both the RFM model and the Schmittlein *et al.* (1987) model and its extensions were developed for noncontractual situations, and, as such, they are briefly only described in this study.

2.6.4.1. Heuristic models: the RFM model

The Recency, Frequency, and Monetary Value (RFM henceforth) model is a heuristic model largely used in customer management.

The RFM approach was firstly proposed by Cullinan in 1978 (Haenlein *et al.*, 2006) and it was extended by Bauer (1988). Nevertheless, Cullinan found the use of these variables to predict future purchase in the 1930s. Also, Gupta *et al.* (2006) and Shih and Liu (2003) argue that this technique has been used for over 30 and 50 years, respectively, in direct marketing.

This approach is based on three dimensions:

- Recency – time elapsed since the last purchase;
- Frequency – number of purchases made within a specific time period;
- Monetary Value – total amount of money spent on purchases from the firm within a specific time period.

RFM is a marketing scoring technique that uses the past purchase behaviour to predict the future customer behaviour. As such, this heuristic is more appropriate for noncontractual situations. The RFM approach has been used to score customers and segment them according to the three dimensions (Colombo and Jiang, 1999; Reinartz and Kumar, 2002, 2003), to allocate resources across customers (Colombo and Jiang, 1999; Reinartz and Kumar, 2002; Shih and Liu, 2003), to target market programs at specific customers (Gupta *et al.*, 2006), to predict partial defection (Buckinx and Van den Poel, 2005), to compute CLV (Dhar and Glazer, 2003) and to estimate customer revenues (Ryals, 2006). Several studies demonstrate that the RFM variables are good predictors for the probability of purchase in the next time period (Etzion *et al.*, 2005).

According to the RFM approach, customers are assumed to be more valuable in the future if they made a larger number of purchases and spent a larger amount of money on purchases from the firm recently than if they made few purchases and spent few money on purchases from the firm some time ago (Dhar and Glazer, 2003). In other words, this technique assumes that the most valuable customers in the future are those who have also been the most valuable in the "recent" past (Schweidel, 2004). But Levin and Zahavi (2001) highlight that recency may work in the reverse way for durable products (*e.g.*, cars), *i.e.*, the likelihood of purchase increases with the rise of recency.

Some advantages of the RFM approach are: it does not require any additional data, it is inexpensive, easy to implement, and easy to understand by managers (Gupta *et al.*, 2006; Haenlein *et al.*, 2006; Shih and Liu, 2003; Villanueva and Hanssens, 2007). Yet, this approach has also some problems. For instance:

- RFM focus on revenues and ignores the costs of acquiring, servicing, and retaining customers; so, the customer profitability is not considered (Dhar and Glazer, 2003; Reinartz and Kumar, 2002; Ryals, 2002a);
- RFM is very simplistic because it only incorporates a limited number of variables (Villanueva and Hanssens, 2007);
- RFM does not take into consideration the effect of the volatility of a customer's past purchasing behaviour on his/her future purchasing behaviour (Dhar and Glazer, 2003);
- RFM ignores the influence of market and macroeconomics variables (Dhar and Glazer, 2003);
- RFM ignores that customers' past purchase behaviour may be the effect of the past marketing activities of the firm (Gupta *et al.*, 2006);

- RFM only allows for the prediction of short-term future behaviour of a customer (Etzion *et al.*, 2005; Fader *et al.*, 2005a).

2.6.4.2. Schmittlein *et al.* (1987) model and its extensions

Schmittlein *et al.* (1987) proposed a Pareto/Negative Binomial Distribution (Pareto/NBD) model that intends to predict the probability of a customer remaining active to the firm, based on recency and frequency. The authors use this model to predict the expected number of transactions in the next period. This model is adequate for noncontractual situations and continuous duration time.

Schmittlein and Peterson (1994) extended the Schmittlein *et al.* (1987) model by incorporating the value of transactions in the model. Fader *et al.* (2005b) proposed a beta-geometric/NBD model which is easier to implement than the model of Schmittlein *et al.* (1987). Fader *et al.* (2004) proposed a beta-geometric/beta-binomial model, which is an extension for discrete time. Fader *et al.* (2006) derive an expression to estimate the Pareto/NBD model parameters using aggregate data. Reinartz and Kumar (2000) implement and extended the approach suggested by Schmittlein and Peterson (1994). They also firstly incorporate the Pareto/NBD model in the CLV. Reinartz and Kumar (2003) replicate the Pareto/NBD model of Reinartz and Kumar (2000), by using the maximum likelihood estimation method. They also suggested a procedure to transform the continuous dependent variable into a dichotomous variable (active/inactive).

2.6.4.3. Binary logistic models and discriminant analysis

In the customer churn context, both binary logistic models and discriminant analysis is used to predict the probability of churn for each customer, and, thus, identify the customers that are likely to cancel a contract with the firm in a pre-specified period, based on some covariates. In this context, the dependent variable is

$$y = \begin{cases} 0 & \textit{the customer does not cancel the contract} \\ 1 & \textit{the customer cancels the contract} \end{cases}$$

Villanueva and Dominique M. Hanssens (2007) present some problems about the use of these methods in the customer churn prediction. Section 3.3 mentions some disadvantages of the binary logistic models in the presence of duration data.

2.6.4.4. Survival models

Continuous survival models are exhaustively described in Chapter 3.

2.6.4.5. Decision trees

In the customer churn context, decision trees may be used to classify the customer into active/inactive. Based on some rules, a kind of tree is constructed. CART and CHAID are some possible algorithms for designing the decision tree. The design of the tree requires a training set.

2.6.4.6. Artificial neural networks

Artificial neural networks (ANN) are non-linear models which can be very useful to model complex situations. A well known problem of these models is that they are considered a "black box". The use of the ANN also requires a training set.

3. CONTINUOUS SURVIVAL ANALYSIS

3.1. Introduction

Survival analysis (also called duration analysis, event history analysis, time-to-event analysis, transition analysis, reliability analysis, or failure time analysis) is a type of econometric method developed to explain and predict the time to event occurrence. The term "survival analysis" has been predominantly used in biomedical sciences, where the variable of interest often is the time to death of patients. Survival analysis is usually called "reliability analysis" or "failure time analysis" in the context of engineering sciences because the focus of engineering people is on the time until machines or electronic components fail. Researchers in the field of social sciences have mainly used the terms "duration analysis" and "event history analysis".

In survival analysis, the event of interest occurs when the individual changes from one state to another one.

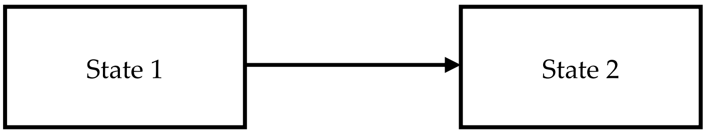
In this study, the term "failure" is sometimes used to denote the event of interest, even though this event may not have necessarily a negative connotation. For instance, in studies about the time of unemployment, the event of interest is that the individual gets a job.

There are two types of survival models, continuous and discrete, depending on whether the event of interest occurs at any instant in time (continuous time) or at discrete time, respectively. Note that survival times are sometimes grouped into discrete intervals of time (*e.g.,* months or years).

In this study, only the continuous survival models are described and used because the event of interest occurs in continuous time. Furthermore, except when mentioned, it is assumed that the population is homogeneous (that is, all the differences between individuals are described by the covariates), all individuals in the study are susceptible of event occurrence, there is only one type of event of interest, the event of interest only occurs once at any individual, and all the covariates are time-invariant. It is also assumed that censoring time and survival time are independent, conditionally on the covariates included in the model.

As such, the most of this study refers to two-state models, which state structure is described in Figure 1.

Figure 1 – State structure of two-state models



Two-state models are Markov models, because "the future depends on the history only through the present" (Putter *et al.*, 2007: 2415).

This chapter is organized as follows. Section 3.2 describes the notions of censoring and truncation. The reasons for the inadequacy of OLS to analyse survival data are exposed in section 3.3. Section 3.4 summarizes the functions of survival time, namely the survival, density, hazard, and integrated hazard functions. In section 3.5 we review the types of survival models, which are divided into three main categories, nonparametric, semi-parametric, and parametric models. Section 3.6 presents a review of literature about frailty (unshared and shared) models. A brief introduction of multiple events models is presented in section 3.7. Lastly, the basis of survival models diagnostics is reviewed in section 3.8.

3.2. Censoring and truncation

One characteristic of survival data is that the information about the survival time of individuals may be incomplete; that is, some information about the survival time of the individual is available, but the exact survival time is unknown (Collet, 1994; Kleinbaum and Klein, 2005; Lee and Wang, 2003). This phenomenon is known as censoring. Censoring occurs when the individual does not fail while under observation, the individual is lost to follow-up during the study, or when the individual fails by different reasons than the event of interest in the study (Cleves *et al.*, 2004; Collet, 1994; Kleinbaum and Klein, 2005; Lee and Wang, 2003).

Let δ be a random variable indicating failure or censoring, defined as $\delta = 1$ if the individual fails during the study period, and $\delta = 0$ otherwise.

According to Cleves *et al.* (2004), Collet (1994), and Hosmer and Lemeshow (1999), there are three types of censoring. Specifically,

- *Right censoring* – when an individual has not experienced the event of interest at the end of the period in analysis (Cleves *et al.*, 2004; Collet, 1994). In this way, the total length of survival time is unknown, and we only know that the completed survival time is of length $T > t$ (Jenkins, 2005). Right censoring is easily dealt both in semi-parametric and parametric models (Cleves *et al.*, 2004).;
- *Interval censoring* – when an individual experiences the event of interest in a known interval of time, but the exact time is unknown (Cleves *et al.*, 2004; Collet, 1994). Interval censoring is easily treated in parametric models but difficult to treat in semi-parametric models.;
- *Left censoring* – when an individual fails before being under observation. Allison (2004) and Collet (1994) designates left censored observations as those whose event of interest occurred before a given unknown time t.

Besides these types of censoring, Klein and Moeschberger (1997) and Allison (2004) mention other type of censoring, which they call random censoring. There is random censoring when "the follow-up stops for reasons that are not under control of the researcher" (*e.g.,* the individual dies, moves away or declines to continue participating in the study) (Allison, 2004: 371). Survival models assume that random censoring is noninformative, that is, "the fact that an individual is censored at a certain point in time does not provide any information about that individual's risk of experiencing the event" (Allison, 2004: 371). Allison (2004) suggests that researchers should try to minimize this type of censoring, because there is no way to correct it. In practice, this type of censoring is treated as right censoring (Allison, 2004).

Survival models assume that the uncensored population represents the independent right-censored sample, which means that censored individuals have the same risk of failure than uncensored individuals (Andersen and Keiding, 2002).

On the other hand, data is truncated when the study design implies a systematic exclusion of observations from the sample (Hosmer and Lemeshow, 1999; Jenkins, 2005; Tableman and Kim, 2004). According to these authors, there are two types of truncation:

- *Right truncation* – the sample only includes those individuals who have experienced the event of interest until a given time. In this way, relatively "long" survival times are systematically not included.;
- *Left-truncation or delayed entry* - the sample excludes all individuals that experienced the event of interest before the delayed entry time in the study.

Cleves *et al.* (2004) refer another type of truncation, denominated interval truncation or gaps, which includes those individuals with gaps, that is, those individuals who are not under observation during a period in the middle of their observation.

Both semi-parametric and parametric models easily handle truncation data, as explained by Cleves *et al.* (2004).

3.3. The inadequacy of OLS to analyse survival data

According to Cleves *et al.* (2004) OLS is not adequate to analyse survival data because OLS assumes that the residuals follow a normal distribution (or, in other words, time conditional on covariates is assumed to be normally distributed), but this assumption is not valid in many situations of survival data. These authors present two examples of events where the assumption of normal distribution of time in unreasonable. They are: (i) an event that has a constant instantaneous risk of failure follows an exponential distribution; and (ii) situations of particular serious surgical procedure, where "many patients die shortly after the surgery, but if they survive, the disease might be expected to return" (p. 2). Even though the normal distribution assumes both positive and non-positive values and survival time is always nonnegative, this inadequacy can be overtaken as suggested by Cleves *et al.* (2004). Linear regression is robust to deviations from normality, but it is not robust to two other characteristics of survival data, which are non-symmetry and non-unimodal (Cleves *et al.*, 2004). Box-Steffensmeier and Jones (2004) and Collet (1994) propose the use of the natural log of survival time to alleviate the skewness problem.

Jenkins (2005) argues that OLS is inadequate to analyse survival data because of (i) the problem of right censoring, and (ii) OLS cannot deal with time-varying covariates. Allison (2004) and Box-Steffensmeier and Jones (2004) also state the problem of traditional regression models to deal with censoring, truncation, and time-varying covariates. Whereas Collet (1994) points out the difficulty of OLS to handle censored observations, Cleves *et al.* (2004) argue that right censoring is not a real problem in linear regression because it can be easily fixed to deal with right censoring (for instance, the software STATA can easily fit these type of models).

Binary dependent variable models, like logit or probit, can be an alternative to OLS that overtakes the censoring and structural modelling problems of OLS (Jenkins, 2005). The dependent variable would be whether or not the event of interest occurs to an individual. But binary dependent variable models have some disadvantages compared to survival models, such as (Allison, 2004; Jenkins, 2005; Kleinbaum and Klein, 2005),

- The survival time of each individual is not considered;
- It does not take into account the exact time at which each person changes the state.

3.4. Functions of survival time

The distribution of survival time is usually described by four main functions: (i) survival function; (ii) density function; (iii) hazard function; and (iv) integrated hazard function. These functions are mathematically equivalent, which means that given one of them, the others can be derived.

Let T be a continuous non-negative random variable, which represents the survival time (measured in minutes, hours, days, years, etc.), t be any specific value of interest for the variable T, and the survival times be independent.

The cumulative density function of T is

$$F(t) = \int_0^t f(x)\,dx = P(T \le t) \qquad \textbf{(1).}$$

3.4.1. Survival function

The survival function is also called survivor function, survivorship function, or reliability function. The survival function is the probability of an individual to survive beyond time t (Collet, 1994), that is

$$S(t) = \int_t^\infty f(x)\,dx = P(T > t) = 1 - F(t) \qquad \textbf{(2).}$$

$S(t)$ is a monotone, nonincreasing function of time t with the following theoretical properties (Cleves *et al.*, 2004; Jenkins, 2005; Klein and Moeschberger, 1997):

$$S(t) = \begin{cases} 1 & for\ t = 0 \\ 0 & for\ t = \infty \end{cases}$$

The graphical representation of $S(t)$ is the survival curve. This curve can be used to determine the median and other percentiles. Examples of survival curves can be found in section 3.5.4 for each parametric model.

3.4.2. Density function

The density function is also called probability density function or unconditional failure rate. The density function of the survival time T is the limit of the probability that an individual fails in a very small (infinitesimal) interval t to $t+\Delta t$ per unit width Δt (Lee and Wang, 2003). Specifically,

$$f(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t+\Delta t)}{\Delta t} = \frac{dF(t)}{dt} = \frac{d}{dt}\left[1-S(t)\right] = -S'(t) \qquad \textbf{(3).}$$

The density function is non-negative and may assume values greater than one, because it is not a set of probabilities (Jenkins, 2005).

The graphical representation of $f(t)$ is the density curve. Some examples of density curves can be found in section 3.5.4, for each parametric survival model.

3.4.3. Hazard function

The hazard function is also called instantaneous failure rate, intensity function (or rate), force of mortality, conditional failure rate, age-specific failure rate (Cleves *et al.*, 2004; Klein and Moeschberger, 1997; Lee and Wang, 2003), risk function, or transition intensity (Andersen and Keiding, 2002). The hazard function is the instantaneous potential per unit time for the event occurrence, given that the individual has survived up to time t (Kleinbaum and Klein, 2005; Tableman and Kim, 2004). That is,

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t < T < t+\Delta t \mid T > t)}{\Delta t} = -\frac{d \ln S(t)}{d\,t} = \frac{f(t)}{1-F(t)} = \frac{f(t)}{S(t)} \qquad \textbf{(4).}$$

The hazard function can vary from zero to infinity, meaning no risk and certainty occurrence of the event of interest at that moment, respectively (Cleves *et al.*, 2004). The hazard function can present a diversity of shapes, such as increasing, decreasing, constant, or even more complicated forms (Lee and Wang, 2003). The graphical representation of $h(t)$ is the hazard curve (see some examples in section 3.5.4, for each parametric survival model).

While the survival function describes de survival experience, the hazard function describes the failure experience. The hazard function is of great interest mainly because (Kleinbaum and Klein, 2005):

- It measures the instantaneous risk whereas the survival function is a cumulative measure;
- It allows to identify a specific parametric model;
- The survival model is usually described by its hazard function.

3.4.4. Integrated hazard function

The integrated hazard function is the total risk of failure accumulated up to time t (Cleves *et al.*, 2004) and is defined as:

$$H(t) = \int_0^t h(x)\,dx = -\int_0^t \frac{1}{S(x)}\left[\frac{d}{dx}S(x)\right]dx = -\ln S(t) \qquad \textbf{(5).}$$

The integrated hazard function has the following theoretical properties:

$$H(t) = \begin{cases} 0 & for\ t = 0 \\ \infty & for\ t = \infty \end{cases}$$

The relationship between the integrated hazard function and the other ones is described as follows:

$$S(t) = \exp(-H(t)) \qquad \textbf{(6),}$$

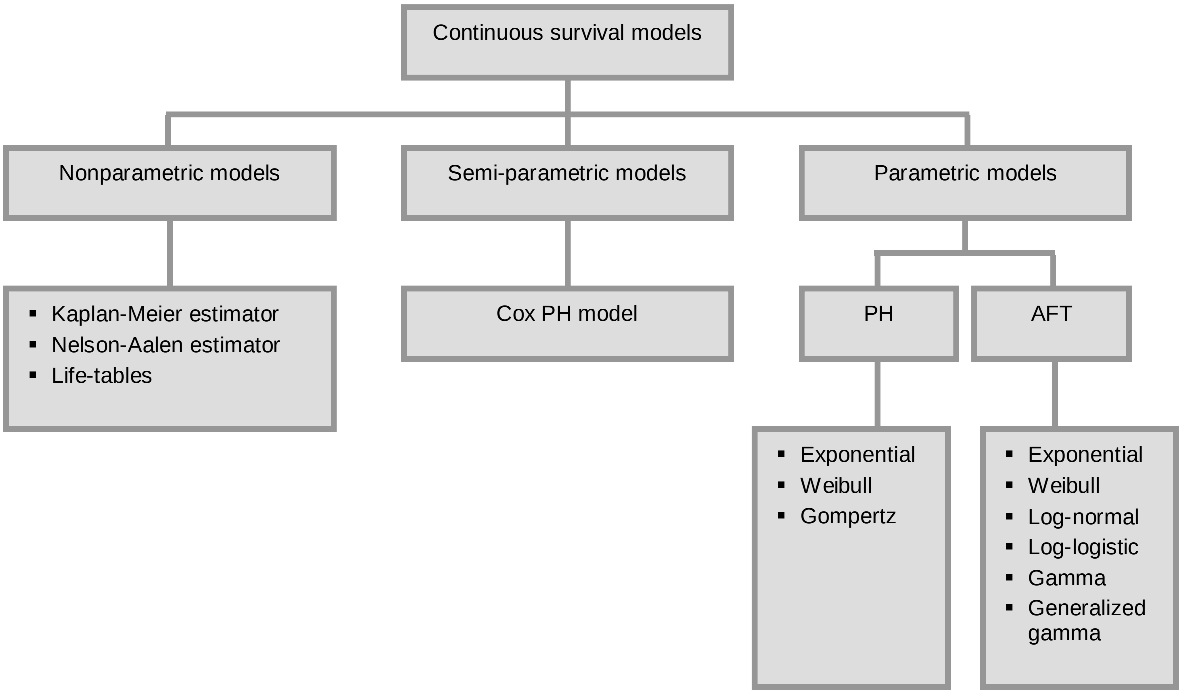
$$F(t) = 1 - \exp(-H(t)) \qquad \textbf{(7),}$$

$$f(t) = h(t)\ \exp(-H(t)) \qquad \textbf{(8).}$$

3.5. Types of models

Continuous survival analysis encompasses several different types of models, which can be grouped into three main categories, namely nonparametric, semi-parametric, and parametric models, as can be seen in Figure 2.

Figure 2 – Types of continuous survival models



3.5.1. Proportional hazards versus accelerated failure time models

Continuous survival models can accommodate both the proportional hazards (PH) and the accelerated failure time (AFT) forms. A PH model is one that satisfies the PH assumption and an AFT model is one that satisfies the AFT assumption.

In PH models the effect of covariates is multiplicative in relation to the hazard, whereas in the AFT models, this effect is multiplicative in relation to the survival time (Allison, 2004; Box-Steffensmeier and Jones, 2004; Kiefer, 1988; Kleinbaum and Klein, 2005).

The Cox semi-parametric model is a PH model. As regards to parametric survival models, exponential and Weibull models can accommodate both the PH and AFT assumptions; Gompertz is a PH model; the others parametric models are AFT models.

3.5.1.1. PH models

PH models are expressed as

$$h(t \mid X) = h_0(t)\exp(\beta' X) \qquad , h_0(t) \geq 0 \qquad \textbf{(9),}$$

where $h_0(t) = h(t \mid X = 0)$ is the baseline hazard, X is the matrix of covariates, β is the vector of the coefficients of the covariates, $\exp(\beta' X)$ is the relative hazard, and $\exp(\beta_i)$ is the hazard ratio of the coefficient of X_i.

PH models assume that the hazard rates of any two individuals are proportional over time (PH assumption) (Allison, 2004; Therneau and Grambsch, 2000). In other words, the PH assumption means that there is a hazard ratio (HR) that is constant and non-negative over time. This assumption can only be satisfied if all covariates are time-invariant. Equation 10 demonstrates this property, as it shows that the hazard ratio does not depend on time.

$$\begin{aligned} HR = \frac{h_i(t \mid X_i)}{h_j(t \mid X_j)} &= \frac{h_0(t)\exp(\beta' X_i)}{h_0(t)\exp(\beta' X_j)} \quad , \quad i \neq j \qquad \textbf{(10).} \\ &= \frac{\exp(\beta' X_i)}{\exp(\beta' X_j)} \\ &= \exp\left[\beta'(X_i - X_j)\right] \end{aligned}$$

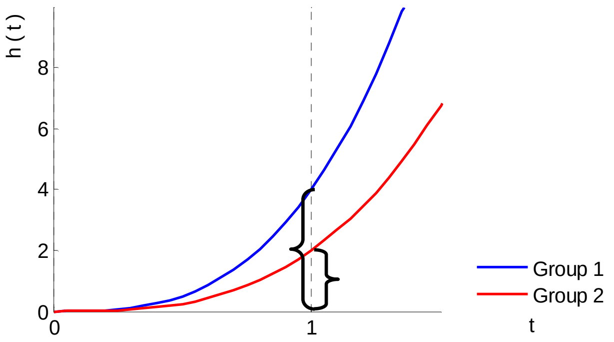
This is equivalent to say that

$$h_i\left(t \mid X_i\right) = HR \times h_j\left(t \mid X_j\right) \quad , \quad i \neq j \qquad \textbf{(11).}$$

As such, the effect of any covariate in the hazard function is constant over time (Allison, 2004; Ata and Sozer, 2007; Hess, 1995). A direct consequence of the PH assumption is that the hazard curves of two distinct individuals or groups of individuals cannot cross (Collet, 1994).

Figure 3 shows the PH assumption by comparing the hazard curves of two groups of individuals. If the hazard ratio is 2, the distance in the vertical line between the t axis and the hazard curve of group 1 is the double of the distance between the t axis and the hazard curve of group 2, at any time t.

Figure 3 - Hazard curves of a PH model ($HR = 2$)



3.5.1.2. AFT models

AFT models can be written as linear models of $\ln(T)$. As such, they are expressed as (Collet, 1994)

$$\ln(T) = \beta' X + \sigma \varepsilon \qquad \textbf{(12),}$$

where ε is a random disturbance with a fixed variance, and σ is a scale parameter that controls the variance of ε. This model assumes that ε is independent of the covariates and that ε_i is independent of ε_j, $i \neq j$ (Allison, 2004).

In order to estimate AFT models by maximum likelihood, the probability distribution of ε must be specified (Allison, 2004; Cleves *et al.*, 2004). There are four distributions that are usually used for ε, which are the normal, the logistic, the extreme value, and the log-gamma distribution (Allison, 2004). The distribution of T depends on the distribution of ε (Allison, 2004; Cleves *et al.*, 2004), as presented in Table 5.

Table 5 – Relationship between the distribution of ε and the distribution of T

Distribution of ε	**Distribution of T**
Extreme value	Weibull or exponential
Normal	Log-normal
Logistic	Log-logistic
Log-gamma	Gamma

Source: Allison (2004: 374)

AFT models assume that there is a constant non-negative acceleration factor that stretches out or shrinks survival times (Collet, 1994). This assumption can be expressed as

$$S_i(t \mid X) = S_0(\psi t) \quad , t \geq 0, \psi \geq 0 \qquad \textbf{(13),}$$

where $S_i(t \mid X)$ is the probability of the individual/group i survives beyond time t, given X, S_0 is the baseline survival, and $\psi = \exp(-\beta' X)$ is the acceleration factor. The implied hazard function is given by

$$h(t \mid X) = \psi\, h_0(\psi t) \qquad \textbf{(14).}$$

It is expected that the event of interest occurs sooner for individuals with $\psi > 1$ and later for individuals with $\psi < 1$ (Cleves *et al.*, 2004; Jenkins, 2005).

An exemplification of this assumption for the case of the relationship between the age of dogs and humans is presented in Kleinbaum and Klein (2005). They state that it is popularly believed that “dogs grow older seven times faster than humans” (p.266), and, as such, the probability that a dog survives more than 10 year equals the probability that a human being survives more than 70 years, that is $S_D(t) = S_H(7\,t)$.

Other interpretation of the AFT model is that the survival time (or the median survival time) of an individual of group i is ψ times the survival time (or the median survival time) of an individual of the reference group (Collet, 1994), that is,

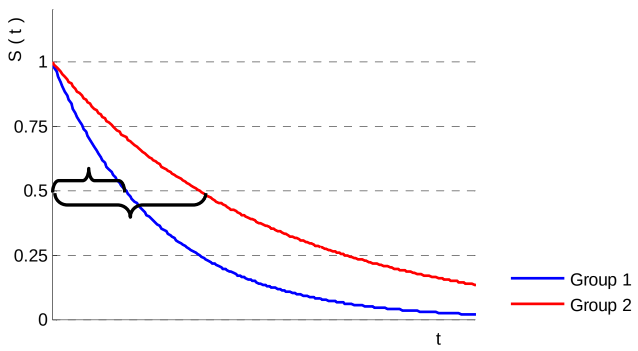
$$Survival\ time_i = \psi \times Survival\ time_0 \qquad \textbf{(15),}$$

and

$$Me(t_i) = \psi \times Me(t_0) \qquad \textbf{(16).}$$

Kleinbaum and Klein (2005) present a graphic that clearly shows the AFT assumption comparing the survival curves of two groups of individuals. Considering an acceleration factor equal to 2, it can be seen that the distance in the horizontal line between the $S(t)$ axis and the survival curve of group 2 is the double of the distance between the $S(t)$ axis and the survival curve of group 1, at any time t (Figure 4).

Figure 4 – Survival curves of an AFT model ($\psi = 2$)



Source: Kleinbaum and Klein (2005: 268) (adapted)

Allison (2004) points out some limitations of the AFT models. He states that the results strongly depend on the selected distribution for ε. On the other hand, he mentions some advantages of the AFT models compared to the Cox-model, such as (i) AFT models handle better left-censoring and irregular interval censoring; and (ii) they easier predict the failure times.

3.5.2. Nonparametric models

Nonparametric models do not make any assumption about the shape of the relevant functions and they do not include any covariate (Blossfeld *et al.*, 2007; Cleves *et al.*, 2004; Collet, 1994). As such, these models are very useful for first exploratory data analysis (Blossfeld *et al.*, 2007). The hazard function is estimated based only on the empirical data of survival time and on the customer status. Covariates can only be considered by stratifying the data into groups and then comparing the survival and hazard functions of the groups. A handicap of the nonparametric models is that they can only handle a small number of groups and only one covariate can be analysed at each time. Moreover, continuous covariates cannot be analysed in nonparametric models, except if they are discretized.

Lee and Wang (2003) point out that nonparametric models are more efficient than parametric models when the adequate theoretical distribution is unknown; but they are less efficient than parametric models when survival time follows a known distribution. In any case, nonparametric models can be useful to choose the theoretical distribution, by the analysis of their survival curves (Cleves *et al.*, 2004; Lee and Wang, 2003).

The (product-limit) Kaplan-Meier method for estimating the survival function, the Nelson-Aalen estimator of the integrated hazard function, and life tables are some examples of nonparametric models.

3.5.2.1. Kaplan-Meier estimator of the survival function

Let $t_1 < t_2 < \cdots < t_j < \cdots < t_k < \infty$ denote the observed failure times and let t_i and t_j be independent $(i \neq j)$. The Kaplan-Meier estimator of survival function (Kaplan and Meier, 1958) is expressed by

$$\hat{S}(t_j) = \prod_{j|t_j<t} \left(1 - \frac{d_j}{n_j}\right) \qquad \textbf{(17),}$$

where d_j is the number of individuals that fail at t_j, n_j is the number of individuals that are at risk of failure immediately prior to t_j (*i.e.*, the number of individuals that survive at least until t_j), which is given by

$$n_j = (m_j + d_j) + (m_{j+1} + d_{j+1}) + \cdots + (m_k + d_k) \qquad \textbf{(18),}$$

and m_j is the number of individuals whose survival time is censored in the interval $[t_j, t_{j+1}[$.

The KM estimates of the survival function can only be determined at uncensored survival times (Jenkins, 2005; Lee and Wang, 2003). Consequently, the survival curve is a step continuous function, starting at $t = 0$ (Collet, 1994).

When there are no censored observations, the estimate of the survival function at time t is the proportion of individuals alive at time t (Allison, 2004; Collet, 1994). Note that when there are tied observations, the Kaplan-Meier estimator assumes that failures occur before censoring (Hosmer and Lemeshow, 1999).

The estimator of the integrated hazard function can be derived from the KM estimator of the survival function, using Equation 6. The estimator of the hazard function cannot be directly derived from the estimator of the integrated hazard function by taking the derivative of the integrated hazard relative to t, because the integrated hazard function is a step function and, as such, its slope is not well-defined (Cleves *et al.*, 2004; Jenkins, 2005). Hougaard (2000) points out that this is the main disadvantage of this kind of models. An alternative to compute the estimator of the hazard function is dividing the survival time into regular intervals of time and determine estimates of the interval hazard function (Collet, 1994; Jenkins, 2005). The hazard function in each interval is given by (Collet, 1994)

$$\hat{h}(t) = \frac{d_j}{n_j \ \tau_j} \quad \textbf{(19),}$$

where $t_j \le t < t_{j+1}$, and $\tau_j = t_{j+1} - t_j$ is the length of the interval j. Note that this equation cannot be used to determine the hazard rate of the interval that starts at the longer failure time, because the interval is open-ended (Collet, 1994).

Another alternative to derive the hazard function is by smoothing the integrated hazard function with the Kernel smoother method, as explained by Klein and Moeschberger (1997) and Cleves *et al.* (2004). The smoothed hazard function is easily derived by using this method (Jenkins, 2005).

Kaplan-Meier is a maximum likelihood estimator, and, as such, it is proved that under certain conditions, its estimates are consistent and asymptotically normal (Lee and Wang, 2003). The KM estimator can easily handle censored and truncated observations (Cleves *et al.*, 2004). Nevertheless, this estimator also presents some drawbacks, as explained by Lee and Wang (2003: 76).

3.5.2.2. Nelson-Aalen estimator of the integrated hazard function

The Nelson-Aalen estimator is the result of a study of Nelson in 1972 and a study of Aalen in 1978 (Cleves *et al.*, 2004). Nelson-Aalen firstly estimates the integrated hazard function and then derives the survival function, which is the inverse process of Kaplan-Meier estimator. The Nelson-Aalen estimator of the integrated hazard function is expressed as

$$\hat{H}(t_j) = \sum_{j|t_j<t} \left(\frac{d_j}{n_j} \right) \qquad \textbf{(20),}$$

and the derived estimate of the survival function (sometimes called Fleming-Harrington estimator) is computed based on Equation 6.

Kaplan-Meier and Nelson-Aalen estimators are asymptotically equivalents (Cleves *et al.*, 2004; Jenkins, 2005). But for small samples, the Nelson-Aalen estimator produces better estimates of the integrated hazard function and the Kaplan-Meier estimator produces better estimates of the survival function (Cleves *et al.*, 2004; Jenkins, 2005).

3.5.2.3. Life-table estimator

There are two main groups of life-tables, population life tables and clinical life tables. The first group can be divided into cohort life table and current life table (Lee and Wang, 2003). But the estimation method is similar for all types of life-tables (Lee and Wang, 2003).

The underlying idea to the life-table estimator is the same as the Kaplan-Meier, but the life-table estimator was developed to suit situations where the number of failures and the number of individuals in the risk set are grouped into intervals of time (Blossfeld *et al.*, 2007; Kalbfleisch and Prentice, 1980; Kiefer, 1988), with similar length or not (Collet, 1994). As survival time is continuous, but they are grouped into intervals, the life-table estimator of the survival function calculates an average estimate for the midpoint of the interval (Jenkins, 2005) and is expressed as (Collet, 1994)

$$\hat{S}(j) = \prod_{k=1}^{j} \left(1 - \frac{d_k}{n_k} \right) \qquad \textbf{(21),}$$

where d_k is the number of individuals that fail within the interval $I_k = [t_k, t_{k+1}[$, $k = 1, 2, , j$, n_k is the adjusted number of individuals in the risk set in the midpoint of the interval I_k, which is expressed as

$$n_k = N_k - \frac{d_k}{2} \quad \textbf{(22),}$$

and N_k is the number of individuals in the risk set at the start of interval.

The estimator of the density function and the hazard function can be derived from the estimator of the survival function. The estimator of the density function is (Lee and Wang, 2003)

$$\hat{f}(j) = \frac{\hat{S}(j) - \hat{S}(j+1)}{t_{j+1} - t_j} \quad \textbf{(23),}$$

and the estimator of the hazard function is given by (Collet, 1994)

$$\hat{h}(j) = \frac{\hat{f}(j)}{\left[\hat{S}(j) + \hat{S}(j+1)\right]/2} \quad \textbf{(24).}$$

3.5.3. Semi-parametric models

3.5.3.1. Cox PH model

The Cox PH model is by far the most popular survival model (Cleves *et al.*, 2004). This model was proposed by Cox (1972) and it is a semi-parametric model, because the baseline hazard is unknown and unparameterized, which means that it is not made any assumption about the shape of the baseline hazard function (Allison, 2004; Kalbfleisch and Prentice, 1980). Thus, even though the shape of the baseline hazard is the same for all individuals (Cleves *et al.*, 2004), the baseline hazard can assume any shape. This can be an advantage of the semi-parametric models compared to parametric models, because if the assumption about the shape of the baseline hazard is wrong, misleading coefficient estimates of the covariates may result (Cleves *et al.*, 2004). On the contrary, when the distribution of the survival time is known, parametric models produce more efficient estimates of the coefficients of covariates (Cleves *et al.*, 2004; Lee and Wang, 2003).

As the baseline hazard is unspecified, the output of the estimates does not contemplate an intercept term (Allison, 2004; Box-Steffensmeier and Jones, 2004; Cleves *et al.*, 2004; Hosmer and Lemeshow, 1999), as it is included in the baseline hazard (Box-Steffensmeier and Jones, 2004; Cleves *et al.*, 2004). The baseline hazard is not directly obtained from the outputs of the estimated Cox model (Cleves *et al.*, 2004). The estimated baseline survival function of the Cox model is equivalent to the Kaplan-Meier estimates and the estimated baseline integrated hazard coincides with the Nelson-Aalen estimates (Cleves *et al.*, 2004).

The Cox model is robust, because it produces estimates that are very similar to those of the correct PH parametric model (Cleves *et al.*, 2004; Kleinbaum and Klein, 2005). In fact, if there are significant differences in the estimated coefficients, the survival data do not follow the specified distribution (Cleves *et al.*, 2004).

The estimation of the semi-parametric model is possible only due to the assumption of PH and the partial likelihood estimation method proposed by Cox (Jenkins, 2005).

The hazard function of the Cox PH model is given by

$$h(t \mid X) = h_0(t)\exp(\beta' X) \qquad , h_0(t) \geq 0 \tag{25}$$

3.5.3.2. Estimation of the Cox model

Cox developed a new estimation method for the Cox PH model, called partial likelihood estimation method (Allison, 2004; Hougaard, 2000; Therneau and Grambsch, 2000). The

Cox model can only be estimated with the partial likelihood method (Jenkins, 2005). Partial likelihood allows estimating the coefficients of the Cox model without imposing any theoretical distribution to the baseline hazard (Allison, 2004; Box-Steffensmeier and Jones, 2004; Menezes, 2004). The coefficients of the model and the baseline hazard are estimated separately; the coefficients are firstly estimated with the partial likelihood method and then the baseline hazard is estimated with other methods as mentioned in section 3.5.3.1 (based on the values of the estimated coefficients) (Collet, 1994).

The partial likelihood method assumes that the time intervals between successive failure times provide no information about the effect of covariates on the hazard function (Collet, 1994). Hence, the hazard function is zero in those time intervals and it only have significant values in failure times (Collet, 1994). It allows the likelihood function to only take into account the order of failure times, ignoring the exact failure times (Collet, 1994; Hougaard, 2000).

According to Kleinbaum and Klein (2005) and Collet (1994) the name “partial” likelihood derives from the fact that it only considers the order of events and does not take into account the exact failure times.

Kalbfleisch and Prentice (1980) point out that in presence of only no-tied failure times Cox proved that the partial likelihood estimation method produces estimates for the parameters with the same properties as MLE. However, in presence of tied failure times, the ordinary partial likelihood proposed by Cox is not a consistent estimator (Kalbfleisch and Prentice, 1980).

3.5.3.2.1. No tied failure times

Let $t_1 < t_2 < \cdots < t_k$ be the ordered failure times. Assuming that there are no ties in failure times, the partial likelihood is given by (Box-Steffensmeier and Jones, 2004)

$$L_p = \prod_{i=1}^{k} L_i \qquad \textbf{(26),}$$

where k is the number of failure times, and L_i denotes the probability of the individual i fails at time t_i, given that he/she is in the risk set at t_i. Equation 26 is equivalent to (Kalbfleisch and Prentice, 1980)

$$L_p = \prod_{i=1}^{k} \left[\frac{\exp(\beta' X_i)}{\sum_{j \in R(t_i)} \exp(\beta' X_j)} \right]^{\delta_i} \qquad \textbf{(27),}$$

where $R(t_i)$ are the individuals in the risk set at time t_i. As censored observations are included in the risk set, they contribute to the denominator but not to the numerator of the partial likelihood function. This implies the following log-likelihood function

$$\ln L_p = \sum_{i=1}^{K} \delta_i \left\{ \beta' X_i - \ln \left[\sum_{j \in R(t_i)} \exp\left(\beta' X_j\right) \right] \right\} \qquad \textbf{(28).}$$

In order to determine the maximum partial likelihood estimates of the model (*i.e.*, the value of the parameters that maximizes the log-likelihood function), the log-likelihood is differentiated with respect to all parameters of the model and then these derivatives are setted equal to 0. Solving the following system of simultaneous equations in order to each parameter, we obtain the maximum partial likelihood estimator for each parameter, under appropriate second order conditions. When several covariates are included in the model, there are several likelihood equations to be solved, as many as the parameters to be estimated. As it is often impossible to analytically solve these equations, iterative procedures are usually used. For instance, Kalbfleisch and Prentice (1980) present a summary of the Newton-Raphson iterative procedure.

3.5.3.2.2. Tied failure times

The partial likelihood method only takes into account the order of the failure time and the exact failure time is not considered (Allison, 2004; Box-Steffensmeier and Jones, 2004;

Jenkins, 2005). So, when there are censored and failure observations at a given failure time, it is assumed that censoring takes place after failures and consequently the risk set at that time includes censored observations (Collet, 1994). However, when there are tied failed observations, the exact ordering of the failure times is impossible to be defined, and, as a consequence, the partial likelihood proposed by Cox cannot be used (Cleves *et al.*, 2004; Collet, 1994). So, some alternatives for this method have been proposed in the literature. They are: (i) the exact partial likelihood method (or average method); (ii) the Breslow method; (iii) the Efron method; and (iv) the exact discrete method (Box-Steffensmeier and Jones, 2004; Cleves *et al.*, 2004; Therneau and Grambsch, 2000).

3.5.3.2.2.1. Exact partial likelihood method

Kalbfleisch and Prentice (1980) proposed a partial likelihood function to handle tied observations that assumes that tied observations are due to an inadequate measurement of the survival time, and the exact order of survival times can be any of the possible $d!$ arrangements of their values (Hosmer and Lemeshow, 1999). This method is designated "exact method" because it accounts for all possible orderings of tied failure times (Box-Steffensmeier and Jones, 2004; Kalbfleisch and Prentice, 1980).

The partial likelihood function of Kalbfleisch and Prentice is given by

$$L_p = \prod_{i=1}^{k} \left[\frac{\exp(\beta' S_i)}{\sum_{j \in R_{d_i}(t_i)} \exp(\beta' S_j)} \right] \qquad \textbf{(29),}$$

where S_i is the sum of all covariates for all individuals that fail at time t_i, $S_j = \sum_{m=1}^{d_i} X_{j_m}$, $j = (j_1, j_2, \cdots, j_{d_i})$, d_i is the number of individuals who fail at time t_i, and $R_{d_i}(t_i)$ is the set of all subsets of d_i items chosen from the risk set at time t_i without replacement.

This method produces reasonably good estimates when there are a large proportion of tied observations in the risk set at time t_i (Kalbfleisch and Prentice, 1980). But if there are many ties at any failure time, this method becomes highly computationally expensive (Kalbfleisch and Prentice, 1980).

3.5.3.2.2.2. Breslow method

This method was proposed by Breslow in 1974 (Box-Steffensmeier and Jones, 2004; Hosmer and Lemeshow, 1999). It is an approximation of the exact partial likelihood method (Hosmer and Lemeshow, 1999; Kalbfleisch and Prentice, 1980). The Breslow method assumes that tied failure times happen in an unknown (and not important) sequence and that the risk set includes all individuals at risk at the failure time (tied and no tied observations) (Box-Steffensmeier and Jones, 2004). The Breslow partial likelihood function is expressed as (Kalbfleisch and Prentice, 1980)

$$L_p = \prod_{i=1}^{k} \frac{\exp\left(\beta' S_i\right)}{\left[\sum_{j \in R(t_i)} \exp\left(\beta' X_j\right)\right]^{d_i}} \qquad \textbf{(30).}$$

This method is adequate when there is a small proportion of tied observations in the risk set at time t_i (Collet, 1994; Klein and Moeschberger, 1997), but when this proportion is large this method may produce large biased estimates of the coefficients of the covariates (Kalbfleisch and Prentice, 1980). Therneau and Grambsch (2000) point out that even though its computation is fast, the Breslow method produces the least accurate estimates. Box-Steffensmeier and Jones (2004) also highlight the simple computation of this method. According to them, this is the most used method to handle tied failure times. Actually, this is the default method in almost all software (Therneau and Grambsch, 2000).

3.5.3.2.2.3. Efron method

Efron method was proposed by Efron in 1977 (Box-Steffensmeier and Jones, 2004; Hosmer and Lemeshow, 1999). This method is also an approximation of the exact partial likelihood method (Cleves *et al.*, 2004; Kalbfleisch and Prentice, 1980). According to the Efron method, the risk set changes depending on the sequence of the tied events (Box-Steffensmeier and Jones, 2004), and, therefore this approximation is more accurate than that proposed by Breslow (Box-Steffensmeier and Jones, 2004; Cleves *et al.*, 2004) but it requires more calculations (Cleves *et al.*, 2004). This method produces quite accurate estimates except if the proportion of ties relative to the size of the risk set is extremely large (Therneau and Grambsch, 2000). When the proportion of tied observations is small, the estimates obtained by this method are quite similar to those of the Breslow method (Klein and Moeschberger, 1997).

The Efron partial likelihood function is given by (Kalbfleisch and Prentice, 1980)

$$L_p = \prod_{i=1}^{k} \left(\frac{\exp\left(\beta' S_i\right)}{\prod_{r=1}^{d_i} \left[\sum_{j \in R(t_i)} \exp\left(\beta' X_j\right) - \frac{(r-1)}{d_i} \; \sum_{j \in D(t_i)} \exp\left(\beta' X_j\right) \right]} \right) \qquad \textbf{(31),}$$

where r is the number of individuals with tied failure times, and $D(t_i)$ is the number of individuals with tied failure times in the risk set at time t_i.

3.5.3.2.2.4. Exact-discrete method

Kalbfleisch and Prentice (1980) suggest the use of this method when ties are frequent. While the above methods assume that time is continuous, the discrete method assumes that time is discrete (Box-Steffensmeier and Jones, 2004).

The exact-discrete method estimates the probability that an individual fails at time t_i, given the composition of the risk set at time t_i (Box-Steffensmeier and Jones, 2004). This method creates groups of individuals based on the risk set at time t_i and the dependent variable is the censoring indicator (Box-Steffensmeier and Jones, 2004). This is equivalent to a conditional logit model (Box-Steffensmeier and Jones, 2004). For more details, see Box-Steffensmeier and Jones (2004).

Hertz-Picciotto and Rockhill (1997) compare three methods to handle tied observations in the partial likelihood: the Breslow, the Efron, and the marginal likelihood of Kalbfleisch and Prentice (1973) methods. In a simulation without censored observations, they found that the most accurate method is that of Efron, namely when the sample is small. Furthermore, they found evidence that whereas the Breslow method tends to underestimate the coefficient of the covariates, the Kalbfleisch and Prentice method tends to produce overestimated coefficients, and that this bias tends to become accentuated as ties increase.

3.5.3.3. The extended Cox model

The ordinary Cox model can be extended by doing stratification, including TVCs or frailty effects. All of these extensions are described below.

3.5.3.3.1. The stratified Cox model

When one or more covariates do not satisfy the PH assumption, they may be used to stratify the Cox model and the other covariates that satisfy this assumption are included in the model (Ata and Sozer, 2007; Blossfeld *et al.*, 1989; Klein and Moeschberger, 1997).

Therneau and Grambsch (2000) present some disadvantages of the stratified Cox model, which are,

- The association between the stratification variable and survival time is not statistically tested because the stratum effects are estimated nonparametrically;
- Whereas stratification using categorical covariates is easy, continuous covariates need to be categorized, which can be a subjective task;
- Using few categories leads to residual bias in the coefficients of the covariates, but a large number of categories leads to efficiency loss;
- Stratified Cox models are less efficient than the ordinary Cox model.

Let $Z_1, Z_2, \cdots, Z_k$ be the covariates that do not satisfy the PH assumption, $X_1, X_2, \cdots, X_p$ be the covariates that satisfy the PH assumption, and Z^* is a variable with $g = 1, 2, \cdots, k^*$ disjoint strata (as many as the number of combinations of the categories of all variables Z_i).

Data are divided based on the values of the variable Z^* and then a stratified Cox model can be estimated. As the stratified variable Z^* is not included in the model, it is not estimated a coefficient for this variable (Therneau and Grambsch, 2000). The hazard function of the stratified Cox model is given by (Kalbfleisch and Prentice, 1980)

$$h_g(t \mid X) = h_{0g}(t)\exp(\beta' X) \qquad \textbf{(32),}$$

where $h_{0g}(t)$ is the baseline hazard of the stratum g. As can be seen in the equation 32, the baseline hazard function varies across groups, but the coefficients of the covariates are constant across groups. This later property of the stratified model is called “no-interaction” assumption (Ata and Sozer, 2007; Kleinbaum and Klein, 2005). In this way, the PH is assumed within each stratum but not across strata (Ata and Sozer, 2007; Collet, 1994; Lee and Wang, 2003). Being an assumption, the “no-interaction” should be tested (Kleinbaum and Klein, 2005).

The hazard function of the interaction model is defined as

$$h_g(t \mid X) = h_{0g}(t)\exp(\beta'_g \; X) \qquad \textbf{(33).}$$

An alternative, but equivalent, hazard function of the interaction model is (Kleinbaum and Klein, 2005)

$$h_g(t \mid X) = h_{0g}(t)\exp\left(\beta' X + \beta_s' X \, Z_s^*\right) \quad \textbf{(34),}$$

where s are dummy variables from Z^*, $s = 1, 2, \cdots, k^* - 1$. If all the coefficients of the interaction terms are statistically significant, the estimates of the stratified model is equivalent to those obtained by estimating separate Cox models for each stratum (Therneau and Grambsch, 2000).

The test of the “no-interaction” assumption is a likelihood ratio test based on a comparison of the log-likelihood of the no-interaction model and the log-likelihood of the interaction model (Ata and Sozer, 2007; Kleinbaum and Klein, 2005). The statistic of the likelihood ratio test is given by (Kleinbaum and Klein, 2005)

$$LR = -2\ln L_R - \left(-2\ln L_F\right) \quad \textbf{(35),}$$

where R is the reduced (no-interaction) model, and F is the full (interaction) model. This statistic follows approximately a chi-squared distribution with p or $p\left(k^* - 1\right)$ degrees of freedom (for the interaction model and for the alternative interaction model, respectively) under the null hypothesis that no-interaction is acceptable (Kleinbaum and Klein, 2005). As such, the interaction model is preferable when there is statistical evidence to reject this null hypothesis (Kleinbaum and Klein, 2005).

Similarly to the ordinary Cox model, the stratified Cox model is estimated using the partial likelihood method. The partial likelihood function of the stratified Cox model is expressed as (Hosmer and Lemeshow, 1999)

$$L_p = \prod_{g=1}^{k^*} L_g \quad \textbf{(36),}$$

where L_g - is the partial likelihood function for the g^{th} stratum.

3.5.3.3.2. The Cox model with time-varying covariates

A time-varying covariate (TVC) is a variable that may vary with time. When the value of TVCs only changes at discrete times t_j, it is assumed that their values remain constant in the k intervals from t_j until t_{j+1} (jump process). Thus, the survival time can be divided into successive intervals of constant covariates. In this situation, the survival function is expressed as (Box-Steffensmeier and Jones, 2004)

$$S\left(t_k\right) = \prod_{j=1}^{k} \mathrm{P}\left(T > t_j \mid T \geq t_{j-1}\right) \qquad \textbf{(37).}$$

TVCs can easily be incorporated into survival models. However, the estimation of these models becomes more complicated (Allison, 2004; Ata and Sozer, 2007; Tableman and Kim, 2004). Allison (2004) mentions two computational methods to estimate survival models with TVCs, and both produce the same results (Allison, 2004). They are:

- *Programming method* – there is a record per individual and the TVC is separated into several covariates, one for each different value of the covariate;
- *Episode splitting method (or counting process method)* – there are multiple records per individual, one for each period during which all covariates are constant. Thus, the TVC only appears once, but its value varies across the records of each individual.

The extended Cox model with TVCs is not a PH model because the PH assumption does not hold with TVCs (Ata and Sozer, 2007; Collet, 1994). The hazard function of the extended Cox model with TVCs is defined as (Blossfeld *et al.*, 1989)

$$h\left(t \mid X, W(t)\right) = h_0(t)\exp\left(\beta' X + \alpha' W(t)\right) \qquad \textbf{(38),}$$

where $W(t)$ is the vector of time-varying covariates. This model assumes that the effect of a given TVC on the survival probability at time t depends only on the value of this TVC at time t and not on the value of the TVC in an earlier time (Kleinbaum and Klein, 2005; Tableman and Kim, 2004). However, the extended Cox model can also include TVCs with lag-time effects (Kleinbaum and Klein, 2005, Tableman and Kim, 2004). In this situation, the hazard function is expressed as (Kleinbaum and Klein, 2005)

$$h\left(t \mid X, W(t)\right) = h_0(t)\exp\left[\beta' X + \alpha' W(t-l)\right] \quad \textbf{(39),}$$

where l is the lag-time for a given TVC.

Using the episode splitting method, the likelihood function of the extended Cox model with TVCs (no tied observations) is expressed as (Hosmer and Lemeshow, 1999)

$$L_p = \prod_{i=1}^{K}\left[\frac{\exp\left[\beta' X_i + \alpha' W_i(t_i)\right]}{\sum_{j\in R(t_i)} \exp\left(\beta' X_j + \alpha' W_j(t_i)\right)}\right]^{\delta_i} \quad \textbf{(40).}$$

When there are tied observations, the methods presented in 3.5.3.2.2 must be used (Box-Steffensmeier and Jones, 2004).

In order to maximize this log-likelihood function, the value of all covariates has to be known at any failure time, which can be difficult to obtain both for internal and external TVC (but mainly for internal ones) (Collet, 1994). When the exact value of the TVC at a given failure time cannot be obtained, an approximation has to be used (Collet, 1994). The alternative approximation methods are (Collet, 1994):

- Use the last known value of the TVC measured before the time at which the value is needed;
- Use the known value of the TVC measured at the nearest time of the time at which the value is needed;

- Use interpolation between the known values of the TVC measured just before and after the time at which the value is needed (note that this method cannot be used with categorical covariates).

3.5.3.3.3. The Cox model with frailty

For reasons of identifiability, the Cox model can only be fitted with shared frailties (and not unshared frailties) (Cleves *et al.*, 2004). Section 3.6.1.2 presents a detailed explanation of the shared frailty models.

3.5.4. Parametric models

Parametric models assume that the data distribution is known and the researcher has to postulate it in advance. However, in survival analysis, time T can follow several known distributions, such as exponential, Weibull, Gompertz, log-normal, log-logistic, gamma, and generalized gamma distribution.

While the Cox model easily accommodates right-censoring data but dealing with left-censoring and interval-censoring data is much more difficult, parametric models easily accommodate all types of censoring (Kleinbaum and Klein, 2005). Like the Cox model, parametric models also handle easily TVC, delayed entry, and gaps (Cleves *et al.*, 2004). But whereas in the Cox model the origin time only means that nobody is at risk prior to that origin time, in parametric models time is very important and the origin time indicates when risk begins accumulating (Cleves *et al.*, 2004). According to Box-Steffensmeier and Jones (2004), parametric models have the following advantages over semi-parametric models: (i) they allow analysing the effect of duration dependence; (ii) it is easy to make predictions beyond the period of analysis (while these predictions are difficult to obtain in the Cox model); (iii) they produce smaller standard errors for the coefficients of the covariates due to efficiency gains derived from the use of MLE (which uses all the information about survival time, instead of the partial likelihood that only uses the order of survival time) (note that for large samples, the standard errors tend to be almost similar).

Moreover, the Cox model may be useful to help selecting the distribution of survival time (Box-Steffensmeier and Jones, 2004), but only if the PH assumption is satisfied.

The parametric models mentioned above are described in the following sections.

3.5.4.1. Exponential model

The exponential model can accommodate both the PH and the AFT forms. These models are similar; they are only written in different ways (Cleves *et al.*, 2004; Kleinbaum and Klein, 2005). Therefore, both models produce the same estimates for the hazard function, survival function and median survival time (Kleinbaum and Klein, 2005).

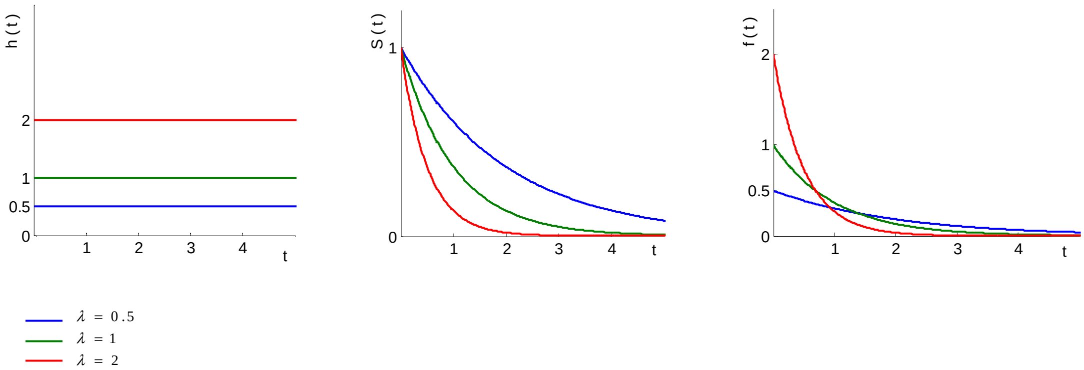
The exponential model is the simplest parametric model because the hazard function of the exponential distribution is constant over time (Collet, 1994); so it does not reflect duration dependence. For this reason, the exponential distribution is often called memoryless (Kiefer, 1988).

Duration dependence occurs when the hazard rate varies with the actual survival time. All the other distributions mentioned above account for duration dependence, and as such, they are much more flexible than the exponential model itself.

This characteristic of the exponential model turns this model very restrictive and sometimes inappropriate for several situations (Box-Steffensmeier and Jones, 2004; Klein and Moeschberger, 1997; Lee and Wang, 2003). Hougaard (2000) points out that this distribution is rarely satisfied. However, Cleves *et al.* (2004) highlight that functions of t can be introduced in the model as covariates, and, in this way, the hazard function is not constant over time.

Figure 5 depicts the hazard function, the survival function, and the density function of the exponential model.

Figure 5 – Exponential model: hazard, survival, and density functions



The hazard function of the exponential model is given by

$$h(t \mid X) = \lambda \quad , \lambda > 0 \tag{41}$$

When $\lambda = 1$, a unit exponential model is obtained (Hougaard, 2000).

In all parametric survival models, the parameter λ is reparameterized in order to incorporate the covariates. Thus, it is considered that $\lambda = \exp(\beta' X)$ in the PH form and that $\lambda = \exp(-\beta' X)$ in the AFT form (Cleves *et al.*, 2004).

The survival function of the exponential model is

$$S(t \mid X) = \exp(-\lambda t), \tag{42}$$

the density function is

$$f(t \mid X) = \lambda \exp(-\lambda t), \tag{43}$$

and the integrated hazard function is given by

$$H(t \mid X) = \lambda t \tag{44}$$

Some descriptive statistics

When survival time follows an exponential distribution, the mean survival time is given by

$$E(T) = \lambda^{-1} \quad \textbf{(45).}$$

The percentile k of the survival time is

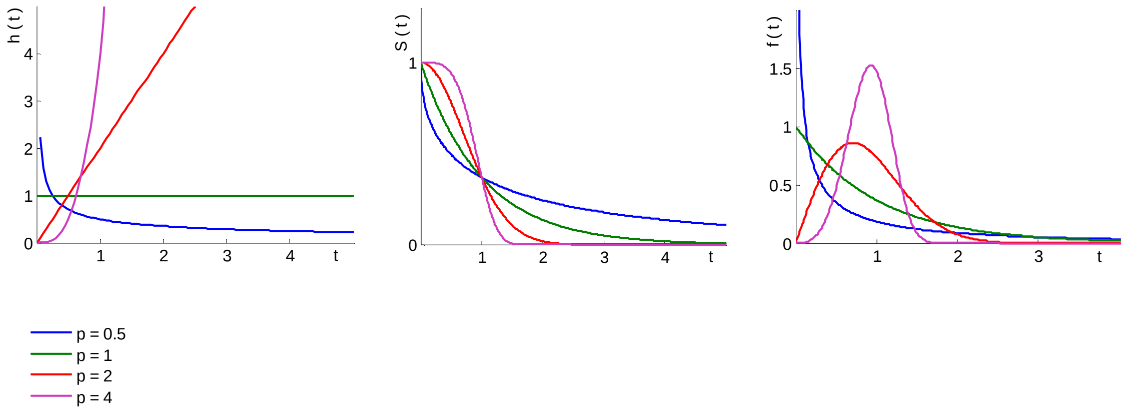
$$t(k) = \lambda^{-1} \ln\left(\frac{100}{100-k}\right) \quad \textbf{(46).}$$

3.5.4.2. Weibull model

The Weibull distribution was proposed by Weibull in 1939 (Lee and Wang, 2003) and it is the most used parametric survival model (Kleinbaum and Klein, 2005). Some areas of previous application are reliability, human disease mortality, and unemployment (Lee and Wang, 2003). The Weibull model can be parameterized as a PH model or an AFT model. Cox and Oakes (1984) proved that the Weibull is the only model where if the PH assumption holds, then the AFT assumption also holds (and vice-versa).

The Weibull distribution has two free parameters, λ and $p = 1/\sigma$, which are the scale and shape parameters, respectively (Box-Steffensmeier and Jones, 2004). The shape parameter affects the shape of the density function, and therefore, the shape of the other functions. The scale parameter stretches or shrinks the distribution. Figure 6 shows the hazard function, the survival function, and the density function of the Weibull model when $\lambda = 1$.

Figure 6 – Weibull model: hazard, survival, and density functions ($\lambda = 1$)



The hazard function of the PH Weibull model is given by (Cleves *et al.*, 2004)

$$h(t \mid X) = \lambda p t^{p-1} \qquad , \lambda > 0,\ p > 0 \tag{47}$$

and the hazard function of the AFT Weibull model is given by (Cleves *et al.*, 2004)

$$h(t \mid X) = \lambda p (\lambda t)^{p-1} \qquad , \lambda > 0,\ p > 0 \tag{48}$$

The hazard function of the Weibull model is monotonically increasing (positive duration dependence) or decreasing (negative duration dependence) as $p > 1$ or $p < 1$, respectively; when $p = 1$, the Weibull distribution reduces to the exponential distribution, which means that the exponential model is a special case of the Weibull model (Kalbfleisch and Prentice, 1980).

The survival function of the PH Weibull model is given by

$$S(t \mid X) = \exp\left(-\lambda t^{p}\right) \tag{49}$$

and the survival function of the AFT Weibull model is

$$S(t \mid X) = \exp\left[-(\lambda t)^{p}\right] \quad \textbf{(50).}$$

The density function of the PH Weibull model is given by

$$f(t \mid X) = \lambda\, p\, t^{p-1} \exp\left(-\lambda t^{p}\right) \quad \textbf{(51),}$$

and the density function of the AFT Weibull model is

$$f(t \mid X) = \lambda\, p(\lambda t)^{p-1} \exp\left[-(\lambda t)^{p}\right] \quad \textbf{(52).}$$

The integrated hazard function of the PH Weibull model is given by

$$H(t \mid X) = \lambda t^{p} \quad \textbf{(53),}$$

and the integrated hazard function of the AFT Weibull model is

$$H(t \mid X) = (\lambda t)^{p} \quad \textbf{(54).}$$

The Weibull model is inappropriate for many situations because it only allows the hazard rate to change in one direction over time (ever-increasing or ever-decreasing hazard rates) (Box-Steffensmeier and Jones, 2004). Even though the Weibull model is more flexible than the exponential model (because it is a function of two parameters, while the exponential model is a function of only one parameter), Weibull is not less flexible than the log-normal and the log-logistic models, because all of them have the same number of parameters (Box-Steffensmeier and Jones, 2004). Nevertheless, in comparison to the Weibull model, the log-normal and the log-logistic models have the advantage of being able to produce non-monotic hazard rates (Box-Steffensmeier and Jones, 2004).

Some descriptive statistics

When survival time follows a Weibull distribution, the mean survival time of the PH model is given by

$$E(T) = \left(\frac{1}{\lambda}\right)^{1/p} \Gamma\left(1+\frac{1}{p}\right) \qquad \textbf{(55),}$$

and the mean survival time of the AFT model is given by

$$E(T) = \frac{1}{\lambda}\Gamma\left(1+\frac{1}{p}\right) \qquad \textbf{(56),}$$

where $\Gamma(p)$ is the well-known gamma function defined as

$$\Gamma(p) = \int_0^\infty x^{p-1} e^{-x} dx = (p-1)! \qquad \textbf{(57).}$$

The percentile k of the survival time of the PH model is given by

$$t(k) = \left[\lambda^{-1} \log\left(\frac{100}{100-k}\right)\right]^{1/p} \qquad \textbf{(58),}$$

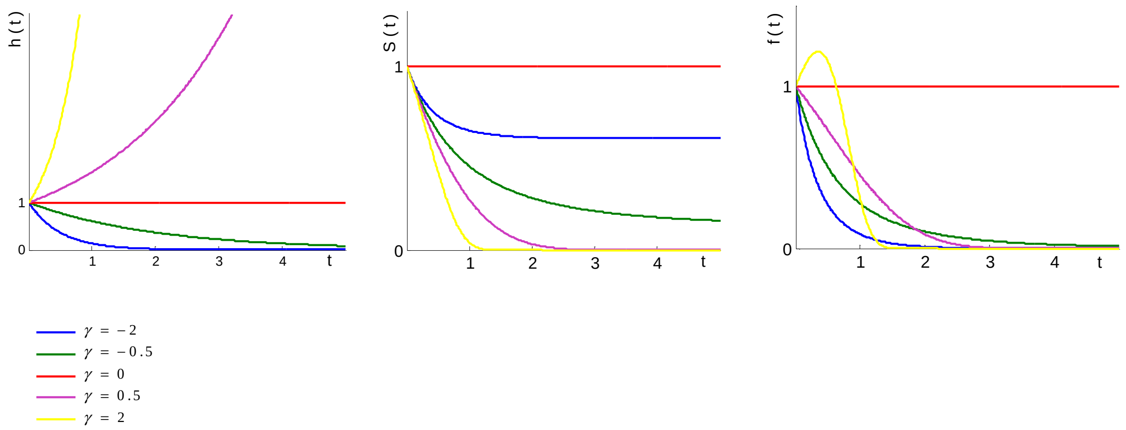
and the percentile k of the survival time of the AFT model is given by

$$t(k) = \lambda^{-1} \log\left(\frac{100}{100-k}\right)^{1/p} \qquad \textbf{(59).}$$

3.5.4.3. Gompertz model

The Gompertz distribution has been widely used by medical researchers and biologists in the study of mortality data (Cleves *et al.*, 2004; Klein and Moeschberger, 1997) and in studies of politics and demographics (Box-Steffensmeier and Jones, 2004). The Gompertz model can only be parameterized as a PH model (Box-Steffensmeier and Jones, 2004; Cleves *et al.*, 2004; Kleinbaum and Klein, 2005). Figure 7 shows the hazard function, the survival function, and the density function of the Gompertz model when $\lambda = 1$.

Figure 7 – Gompertz model: hazard, survival, and density functions ($\lambda = 1$)



The hazard function of the Gompertz distribution is given by

$$h(t \mid X) = \lambda \exp(\gamma t) \qquad , \lambda > 0 \qquad \textbf{(60).}$$

The Gompertz distribution has two parameters, λ (scale parameter) and γ (shape parameter) (Cleves *et al.*, 2004). When $\gamma > 0$, the hazard function increases over time (starting at λ); when $\gamma < 0$, the hazard function falls with time (starting at λ); and when $\gamma = 0$ the hazard is flat over time and the model reduces to an exponential model (Lee and Wang, 2003). Klein and Moeschberger (2003) suggest that γ can be restricted to assume only positive values, because when $\gamma < 0$, the survival function will never be zero as $t \to \infty$, which means that there is a probability of living forever.

The survival function of the Gompertz model is expressed as

$$S(t \mid X) = \exp\left\{\frac{\lambda}{\gamma}\left[1-\exp(\gamma t)\right]\right\} \qquad \textbf{(61),}$$

the density function is

$$f(t \mid X) = \lambda \exp(\gamma t) \exp\left\{\frac{\lambda}{\gamma}\left[1-\exp(\gamma t)\right]\right\}, \qquad \textbf{(62),}$$

and the integrated hazard function is

$$H(t \mid X) = \frac{\lambda}{\gamma}\left[\exp(\gamma t)-1\right] \qquad \textbf{(63).}$$

Some descriptive statistics

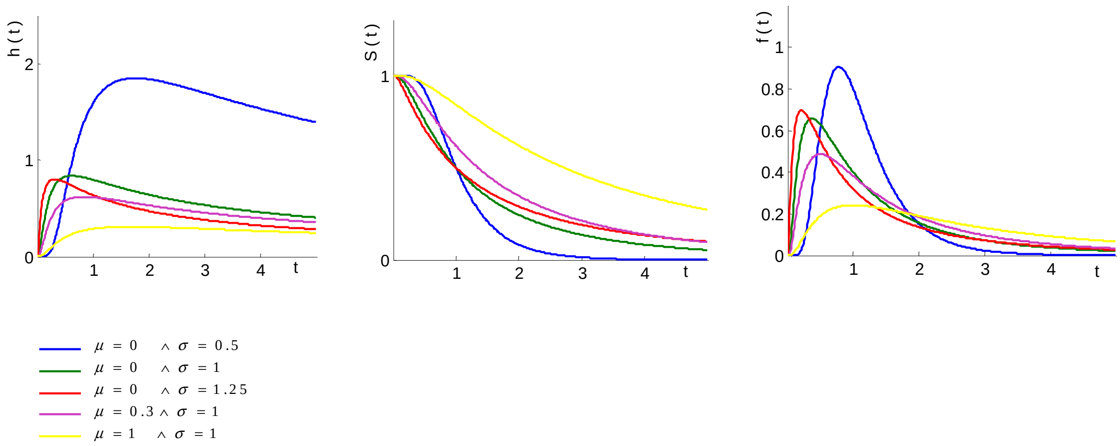
There is no closed-form expression for the mean (Jenkins, 2005). The percentile k of the survival time is given by

$$t(k) = \frac{1}{\gamma}\ln\left[1+\frac{\gamma \ln(100/k)}{\lambda}\right] \qquad \textbf{(64).}$$

3.5.4.4. Log-normal model

The log-normal distribution was proposed by McAlister in 1879, and it has been used in survival models in a large number of areas, such as economics and medicine (Lee and Wang, 2003). The log-normal model is an AFT model. Figure 8 illustrates the hazard function, the survival function, and the density function of the log-normal model when $\lambda = 1$.

Figure 8 – Log-normal model: hazard, survival, and density functions ($\lambda = 1$)



The hazard function of the log-normal model is given by (Lee and Wang, 2003)

$$h(t \mid X) = \frac{\frac{1}{\sigma\sqrt{2\pi}} t^{-1} \exp\left\{-\frac{1}{2}\left[\frac{\ln(t\lambda)}{\sigma}\right]^2\right\}}{1 - \Phi\left[\frac{\ln(t\lambda)}{\sigma}\right]} \qquad , \sigma > 0 \qquad \textbf{(65),}$$

where Φ is the integrated distribution function for the standard normal distribution, and $\sigma = p^{-1}$. This hazard function is characterized by two scale parameters, $\mu = \beta' X$ and σ. The hazard of the log-normal model is hump-shaped (Klein and Moeschberger, 1997), because it rises from zero to a maximum (which is close to the median) and then falls to zero as $t \to \infty$ (Kalbfleisch and Prentice, 1980; Klein and Moeschberger, 1997). This function is positively skewed, and the skewness is greater as greater is the value of σ (Lee and Wang, 2003). One advantage of this model is that the hazard is not monotonic (Cleves *et al.*, 2004). The parameter σ indicates how quickly the hazard rate rises to its peak. Thus, when σ is large, the hazard function reaches its peak very quickly and then drops (Box-Steffensmeier and Jones, 2004).

The survival function of the log-normal model is given by

$$S(t \mid X) = 1 - \Phi\left[\frac{\ln(t\lambda)}{\sigma}\right] \qquad (66),$$

the density function is

$$f(t \mid X) = \frac{1}{\sigma\sqrt{2\pi}} t^{-1} \exp\left\{-\frac{1}{2}\left[\frac{\ln(t\lambda)}{\sigma}\right]^2\right\} \qquad (67),$$

and the integrated hazard function is

$$H(t \mid X) = -\log\left\{1 - \Phi\left[\frac{\ln(t\lambda)}{\sigma}\right]\right\} \qquad (68).$$

Some descriptive statistics

The mean survival time of the log-normal model is given by

$$E(T) = \lambda^{-1} \exp\left(\frac{\sigma^2}{2}\right) \qquad (69).$$

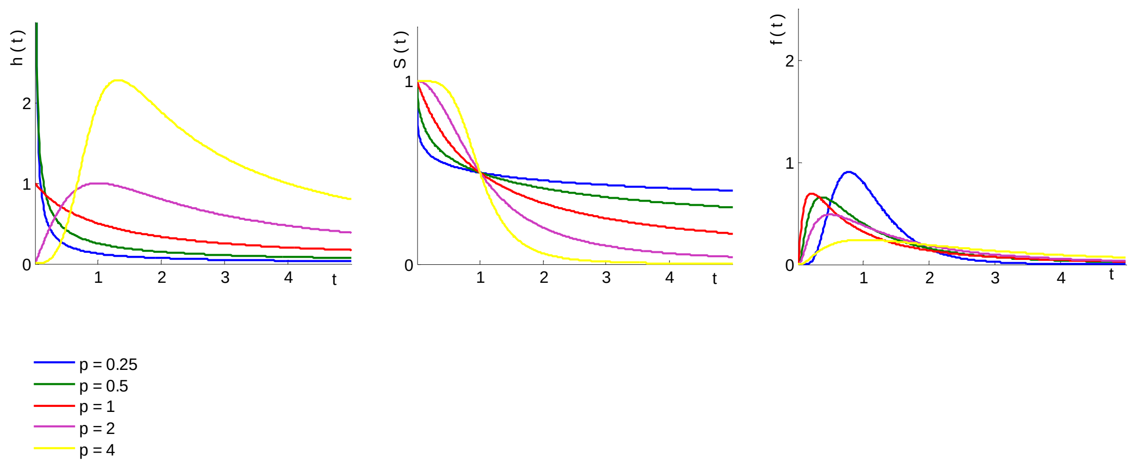
The percentile k of the survival time is given by

$$t(k) = \lambda^{-1} \exp\left(\frac{-\Phi^{-1} - \frac{k}{100}}{p}\right) \qquad (70).$$

3.5.4.5. Log-logistic model

Despite having simpler mathematical expressions of the hazard and survival functions, the shape of the log-logistic distribution is very similar to the log-normal, except on the extreme tail of the distribution (Kalbfleisch and Prentice, 1980; Klein and Moeschberger, 1997; Tableman and Kim, 2004), because the log-logistic distribution has heavier tails than the log-normal distribution (Kalbfleisch and Prentice, 1980). Figure 9 shows the hazard function, the survival function, and the density function of the log-logistic model when $\lambda = 1$.

Figure 9 – Log-logistic model: hazard, survival, and density functions ($\lambda = 1$)



The log-logistic distribution produces a hazard function that can be non-monotonic and unimodal. The hazard function is given by

$$h(t \mid X) = \frac{\lambda p (\lambda t)^{p-1}}{1 + (\lambda t)^{p}} \qquad , \lambda > 0, \ p > 0 \qquad \textbf{(71).}$$

As can be seen, this function has two parameters, λ and $p = 1/\gamma$, which are the scale and the shape parameters, respectively (Box-Steffensmeier and Jones, 2004). When , $p > 1$ the hazard first increases from the origin until reaches a maximum at time $t = \lambda^{-1}(p-1)^{1/p}$ and then falls to zero as $t \to \infty$ (similar behaviour to the log-normal model); when $p < 1$, the hazard starts at infinity and then monotonically decreases with time (similar to the Weibull model); and when $p=1$, the hazards is monotonically decreasing (Kalbfleisch and Prentice, 1980). As such, the log-logistic can be a good alternative for modelling the survival time of patients of heart transplantation, who have an increasing risk of death over the first days after the transplant, and then the risk falls (Collet, 1994).

The survival function of the log-logistic model is

$$S(t \mid X)=\frac{1}{1+(\lambda t)^{p}} \qquad \textbf{(72),}$$

the density function is

$$f(t \mid X)=\frac{\lambda p(\lambda t)^{p-1}}{\left[1+(\lambda t)^{p}\right]^{2}} \qquad \textbf{(73),}$$

and the integrated hazard function is

$$H(t \mid X)=\log\left[1+(\lambda t)^{p}\right] \qquad \textbf{(74).}$$

<u>Some descriptive statistics</u>

The mean survival time of the log-logistic model is given by

$$E(T)=\frac{1}{\lambda p}\times\frac{\pi}{\sin\left(p^{-1}\pi\right)} \qquad , p>1 \qquad \textbf{(75).}$$

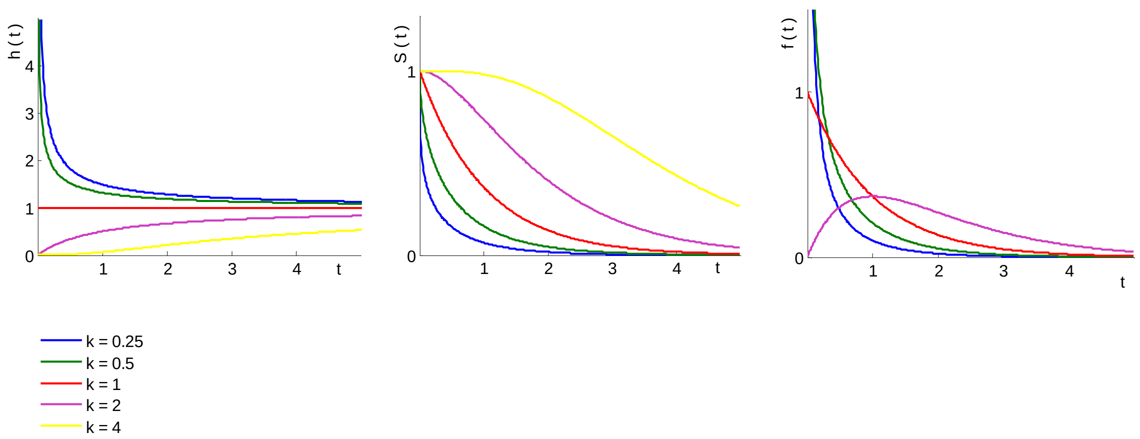
The percentile k of the survival time is given by

$$t(k)=\lambda^{-1}\left(\frac{100-k}{k}\right)^{1/p} \qquad \textbf{(76).}$$

3.5.4.6. Gamma model

The gamma model is an AFT model. The gamma distribution has been applied in studies about industrial reliability and human survival (Lee and Wang, 2003). Figure 10 depicts the hazard function, the survival function, and the density function of the gamma model when $\lambda = 1$.

Figure 10 – Gamma model: hazard, survival, and density functions ($\lambda = 1$)



The hazard function of this model is expressed as

$$h(t \mid X) = \frac{\lambda(\lambda t)^{k-1}\exp(-\lambda t)}{\int_t^{\infty} \lambda(\lambda t)^{k-1}\exp(-\lambda t)\, dt} \qquad , \lambda > 0, k > 0 \qquad \textbf{(77).}$$

Both the Weibull and gamma distributions are generalizations of the exponential distribution, and, as such, Weibull and gamma models generate very similar hazard functions.

The gamma model has two free parameters, λ and k, which are the scale and shape parameters, respectively (Klein and Moeschberger, 1997). When $k < 1$, the hazard rate falls monotonically from infinity to λ as $t \to \infty$; when $k > 1$, the hazard rate rises monotonically from 0 to λ as $t \to \infty$; when $k = 1$, the hazard rate is constant over time, and we have an exponential model (Kalbfleisch and Prentice, 1980; Klein and Moeschberger, 1997).

The survival function of the gamma model is defined as

$$S(t \mid X) = \frac{\int_t^{\infty} \lambda(\lambda t)^{k-1} \exp(-\lambda t)\, dt}{\Gamma(k)} = 1 - I(\lambda t, k) \qquad , \lambda > 0, k > 0 \qquad \textbf{(78),}$$

where $\Gamma(k)$ is defined as presented in Equation 57 and I is the incomplete gamma function. The density function is

$$f(t \mid X) = \frac{\lambda(\lambda t)^{k-1} \exp(-\lambda t)}{\Gamma(k)} \qquad , \lambda > 0, k > 0 \qquad \textbf{(79).}$$

Some descriptive statistics

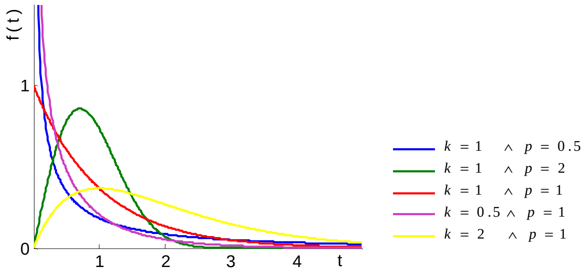
The mean survival time of the gamma model is given by

$$E(T) = \frac{k}{\lambda} \qquad \textbf{(80).}$$

3.5.4.7. Generalised gamma model

The generalized gamma distribution is the most flexible parameterization (Kleinbaum and Klein, 2005), because it is characterized by three free parameters, a scale parameter (λ) and two shape parameters (p and k), and thus, it allows for several possible shapes of the hazard function (Box-Steffensmeier and Jones, 2004; Cleves *et al.*, 2004). Figure 11 shows the density function of the generalized gamma model.

Figure 11 – Generalized gamma model: density function ($\lambda = 1$)



The density function of the generalized gamma distribution is (Kalbfleisch and Prentice, 1980)

$$f(t \mid X) = \frac{\lambda\, p(\lambda t)^{pk-1} \exp\left(-(\lambda t)^{p}\right)}{\Gamma(k)} \qquad , \lambda > 0,\ p > 0,\ k > 0 \qquad \textbf{(81),}$$

where $\Gamma(k)$ is defined as presented in Equation 57. The term "generalized" gamma is due to the fact that other distributions are implied from this distribution, given some specific values of the shape parameters (Box-Steffensmeier and Jones, 2004). So, when $k = 1$ this model reduces to the Weibull distribution; when $k = p = 1$ it reduces to the exponential distribution; when $k = 0$ the log-normal results; and when $p = 1$, we have the standard Gamma distribution (Kalbfleisch and Prentice, 1980). In this way, the generalized gamma model is usually used to test the model specification among the nested-models (exponential, Weibull, log-normal and gamma) (Cleves *et al.*, 2004; Klein and Moeschberger, 1997), as will be presented in section 3.5.4.8.

3.5.4.8. Choosing among parametric models

Choosing the parametric distribution of survival time is a very important (but difficult) task, because if an incorrect distribution is used, the generated estimates are misleading (Box-Steffensmeier and Jones, 2004). When the researcher has some knowledge about the distribution of the survival time, parametric models are the most appropriate to be used because they produce the most efficient estimates of the coefficients of covariates (Cleves *et al.*, 2004) and the estimates are more precise (*i.e.*, they produce estimates with smaller standard errors) (Bradburn *et al.*, 2003). Cleves *et al.* (2004) suggest the use of a semi-parametric model when researchers do not have an idea about the shape of the distribution of survival time and the PH assumption is satisfied.

Even though the parametric distribution should ideally be chosen based on theory (Box-Steffensmeier and Jones, 2004; Cleves *et al.*, 2004), Box-Steffensmeier and Jones (2004) are reluctant to believe that social sciences develop such theory.

Cleves *et al.* (2004) suggest two different statistical strategies to decide which parametric model is more appropriate for the data; one strategy is only suitable for the nested models and the other can be used to decide among nested and non-nested models.

According to these authors, in order to choose among the parametric nested models, a generalized gamma model may be fitted and then the following null hypothesis are tested:

- $H_{0_i}: k = 1 \wedge p = 1$ (survival time follows a exponential distribution)
- $H_{0_{ii}}: k = 1$ (survival time follows a Weibull distribution)
- $H_{0_{iii}}: k = 0$ (survival time follows a log-normal distribution)
- $H_{0_{iv}}: p = 1$ (survival time follows a gamma distribution)

These null hypotheses can be tested using two asymptotically equivalent tests, the likelihood-ratio and the Wald test (Cleves *et al.*, 2004; Lee and Wang, 2003). But under the presence of sample weights or robust estimates of the variance-covariance matrix of the parameters, only the Wald test can be used (Cleves *et al.*, 2004). The statistics of these tests are presented in Lee and Wang (2003).

Neither the likelihood ratio test nor the Wald test can be used to compare non-nested models (Box-Steffensmeier and Jones, 2004; Klein and Moeschberger, 1997). The alternative is a comparison of the Akaike Information Criterion (AIC) proposed by Akaike (1974) or the Bayesian Information Criterion (BIC) proposed by Schwarz (1978).

AIC and BIC may be computed for each model to compare parametric models, either nested or non-nested. Both criteria penalise the log-likelihood of the model for each parameter that is estimated, which means that they compensate the more parsimonious model (Schwarz, 1978). AIC is expressed as

$$AIC = -2\ln L + 2(k + c) \qquad \textbf{(82),}$$

where k is the number of covariates in the model, and c is the number of model-specific distributional parameters. The model with the lowest AIC should be preferred (Akaike, 1974).

The BIC is expressed as

$$BIC = -2\ln L + (k + c)\ln n \qquad \textbf{(83),}$$

where n is the total number of observations. Similarly to AIC, the best model is that with the lowest BIC. In this study, the decision about the parametric model that best fits the data is based on AIC.

The analysis of the goodness-of-fit of the models based on the Cox-Snell residuals also allows choosing among the parametric models (Box-Steffensmeier and Jones, 2004).

3.5.4.9. Estimation of parametric models

All parametric models can be estimated by maximum likelihood (MLE) (Blossfeld *et al.*, 2007; Box-Steffensmeier and Jones, 2004). MLE easily handles censored observations (Allison, 2004). Under certain regularity conditions, MLE generates estimates for the coefficients that are asymptotically unbiased and asymptotically efficient.

Let us suppose we have n observations with $t_1, t_2, \cdots t_n$ survival times, observed survival times are independent (conditional on any covariates), survival time may be censored or uncensored, censored observations are non-informative, observations are independent and identically distributed (i.i.d.).

The likelihood function for survival data depends on two components, the density function and the survival function, and is given by (Blossfeld *et al.*, 1989)

$$L = \prod_{i=1}^{n} \left[f\left(t_i \mid X_i\right)\right]^{\delta_i} \left[S\left(t_i \mid X_i\right)\right]^{1-\delta_i} \qquad \textbf{(84),}$$

where δ_i is the censoring indicator ($i = 1$ if the individual fails during the observation period; $i = 0$, otherwise). Thus, censored observations only influence the likelihood function through the survival function and non-censored observations through the density function (Blossfeld *et al.*, 2007). The implied log-likelihood function is given by

$$\ln L = \sum_{i=1}^{n} \delta_i \ln\left[f\left(t_i \mid X_i\right)\right] + \sum_{i=1}^{n} \left(1-\delta_i\right) \ln\left[S\left(t_i \mid X_i\right)\right] \qquad \textbf{(85).}$$

When TVCs are incorporated into parametric models, the likelihood function is expressed as (using the episode splitting method) (Box-Steffensmeier and Jones, 2004)

$$L = \prod_{i=1}^{k} \left[f\left(t \mid X, W(t)\right)\right]^{\delta_i} \left[S\left(t \mid X, W(t)\right)\right]^{1-\delta_i} \qquad \textbf{(86),}$$

where k is the number of intervals on which TVCs are divided.

3.6. Frailty models

All survival models mentioned above assume that the differences between individuals are all included in the covariates, that is, individuals are similar in all other aspects not measured by the covariates. But that may not be always absolutely true, because of the existence of omitted covariates (Allison, 2004; Blossfeld *et al.*, 1989; Hosmer and Lemeshow, 1999; Karim, 2008). This may happen due to several reasons, such as (i) it is impossible to include all variables that distinguish the individuals, (ii) some important variables are unknown, and (iii) some variables are immeasurable (Blossfeld *et al.*, 2007; Box-Steffensmeier and Jones, 2004; Hougaard, 2000; Karim, 2008; Wienke, 2003).

In these situations, unobserved heterogeneity effects should be included in the model (Blossfeld *et al.*, 1989; Karim, 2008). Unobserved heterogeneity means that some individuals (or groups) are more frail (that is, they are more susceptible to fail) than others for unknown or unmeasured reasons. When these effects are important but omitted in the model, the following consequences may happen:

- The model will over-estimate the degree of negative duration dependence in the hazard function; or, in other words, there is a tendency for the hazard function to decrease faster over time or increase slowly (Aalen and Gjessing, 2005; Allison, 2004; Blossfeld *et al.*, 1989; Jenkins, 2005; Wienke, 2003);
- The coefficients of the covariates will be underestimated (Allison, 2004; Blossfeld *et al.*, 1989; Box-Steffensmeier and Jones, 2004; Henderson and Oman, 1999; Jenkins, 2005; Karim, 2008; Wienke, 2003; Yashin *et al.*, 2001). If $|\beta_{omitted}|$ is large, the bias will be large; otherwise, the bias will be small (Struthers and Kalbfleisch, 1986). Henderson and Oman (1999) found that the extent of bias depends on the variability of the frailties and on the postulated frailty distribution. They also found that in presence of censored observations, the bias is diminished.

Zorn (2000) distinguishes between "true" duration dependence (or state dependence) and "spurious" duration dependence (unobserved heterogeneity).

There are two types of approaches to account for unobserved heterogeneity in survival models, which are the robust estimation and the estimation of frailty models (Box-Steffensmeier and Zorn, 1999). Some controversy exists about which approach is more adequate to this phenomenon and also in presence of multiple events (Box-Steffensmeier and De Boef, 2002). In this study, only frailty models are described and used.

Even though frailty models were firstly introduced by Clayton (1978), the term "frailty" was introduced by Vaupel *et al.* (1979). Frailty models are also called conditional or mixture models (Box-Steffensmeier and Zorn, 1999). It is expected that frailty models are more efficient than the variance-corrected models (also designated as variance-corrected models), if the frailty distribution had been correctly specified (Lin, 1994).

Frailty models have been mainly studied in the context of PH models (Duchateau and Janssen, 2007). Yashin *et al.* (2001) discuss some myths about frailty models, which are mainly due to the ignorance or the misunderstood of the limitations of the model.

Frailties are latent variables that have a multiplicative scale effect on the hazard function. It is assumed that:

- Frailties are random positive values with mean 1 (assumed for purposes of model identificability) and finite variance θ (Blossfeld *et al.*, 1989; Gutierrez, 2002);
- Frailty is constant over time (Blossfeld *et al.*, 2007; Hougaard, 2000),
- Omitted covariates are independent of survival time, of the covariates included in the model (Blossfeld *et al.*, 1989; Yashin *et al.*, 2001), and of any censoring (Hosmer and Lemeshow, 1999).

Frailties are not directly estimated from the data but its variance θ is (Cleves *et al.*, 2004; Gutierrez, 2002).

In random frailty models, there are two sources of variability of the survival times, which are that explained by the included covariates in the model and the frailty term.

3.6.1. Types of frailty models

There are two types of frailty models, (i) univariate survival models, and (ii) multivariate survival models. They are described below.

3.6.1.1. Univariate survival models

Univariate survival models (also designated as unshared frailty models) account for individual unobserved heterogeneity, which means that each individual has its own frailty. Thus, the survival time of an individual is assumed to be independent of the survival time of the other individuals (Karim, 2008; Kleinbaum and Klein, 2005).

The underlying idea in univariate survival models is that individuals are not homogeneous and, as such, they have different frailties. Thus, the most frail will fail earlier and the proportion of robust individuals in the population will increase over time. This is known as the selection process of robust individuals. As the population survival function is the weighted average of the survival function of the several groups that may exist in the population, the selection process will influence the population survival time in a upward trend, because the proportion of robust individuals tends to increase over time. As such, the population hazard function will fall over time. This trend may not reflect the individual hazards, but simply the selection process. Blossfeld *et al.* (1989, 2007) present some examples of this effect in an analytical and graphical way.

The individual (conditional) hazard function of the unshared frailty model is expressed as (Blossfeld *et al.*, 1989)

$$h\left(t \mid X,\alpha\right)=\alpha\, h\left(t \mid X\right) \qquad \textbf{(87),}$$

where α is the unshared-frailty, *i.e.*, the individual unobserved heterogeneity.

The existence of individual frailties can be seen as a misspecification problem because if the unmeasured variables were included in the model, the frailty effect α would be 1 with probability 1 and the ordinary (no-frailty) survival model would result (Box-Steffensmeier and Jones, 2004; Karim, 2008).

Individuals with above-average values of α fail faster (that is, they are more frail) due to unmeasured variables and individuals with below-average values of α fail slowly (that is, they are less frail) due to unmeasured variables (Gutierrez, 2002; Hosmer and Lemeshow, 1999; Jenkins, 2005; Karim, 2008). In other words, if $\alpha > 1$, the frailty effect will increase the individual hazard function and if $\alpha < 1$, the effect will be the opposite (Cleves *et al.*, 2004).

If $\theta = 0$, the model reduces to the ordinary (no-frailty) survival model (Box-Steffensmeier and Jones, 2004). This hypothesis can be tested in order to evaluate the existence of individual unobserved heterogeneity (Box-Steffensmeier and Jones, 2004).

The relationship between the frailty survival function and the no-frailty survival function is expressed as (Gutierrez, 2002)

$$S(t \mid X, \alpha) = [S(t \mid X)]^{\alpha} \qquad \textbf{(88).}$$

So, in order to obtain the population (or unconditional) survival function, the frailty α have to be integrated out by specifying a theoretical distribution with probability density function $g(\alpha)$ for the random variable α, whose functional form is defined by only a few parameters (Blossfeld *et al.*, 1989; Cleves *et al.*, 2004; Gutierrez, 2002). As such, frailty models are mixture models (Hougaard, 2000), because it is assumed a distribution for the hazard function and a distribution for frailties (Cleves *et al.*, 2004).

The population (or unconditional) survival function is defined as (Gutierrez, 2002)

$$S_{\theta}(t \mid X) = \int_{0}^{\infty} \left[S(t \mid X) \right]^{\alpha} g(\alpha)\, d\alpha \qquad \textbf{(89).}$$

where S_{θ} is the population (or unconditional) survival function, that is the survival function that represents a population average, and S is the individual (or conditional) survival function, expressed as $S(t \mid X, \alpha)$.

The population (or unconditional) density function is

$$f_{\theta}(t \mid X) = \int_{0}^{\infty} f(t \mid X, \alpha)\, g(\alpha)\, d\alpha \qquad \textbf{(90).}$$

The population (or unconditional) survival function only depends on the free parameters of the distribution of T, on the effects of the covariates included in the model, on the random coefficient θ, and on the assumed frailty distribution (Gutierrez, 2002).

Considering the relationships between the survival and the hazard functions, the population hazard function (or unconditional on α) is defined as

$$h_{\theta}(t \mid X) = -\frac{d}{dt} S_{\theta}(t \mid X, \theta)\left[S_{\theta}(t \mid X, \theta)\right]^{-1} \qquad \textbf{(91).}$$

which is equivalent to

$$h_{\theta}(t \mid X) = h(t \mid X)\, E(\alpha \mid T > t) \qquad \textbf{(92),}$$

which means that the population hazard function is the average hazard over the survival individuals at any given time (Hougaard, 1995). Thus, in frailty models, the population hazard function is different from the individual hazard function (it is possible that the population hazard decreases while all the individual hazards rise). In no-frailty models, the population and the individual hazard functions are equivalent, because it is assumed that all the individuals are identical in all aspects not measured by the covariates included in the model.

Unshared frailty models are estimated using the MLE method (Blossfeld *et al.*, 1989; Cleves *et al.*, 2004; Kleinbaum and Klein, 2005).

3.6.1.2. Multivariate survival models

In multivariate survival models, the population/sample is divided into some groups of individuals and the individuals of the same group are assumed to be correlated (Cleves *et al.*, 2004; Karim, 2008; Kleinbaum and Klein, 2005).

Shared frailty models account for unobserved heterogeneity of independent groups of individuals (Hougaard, 2000), which means that each group of individuals has its own frailty that may be different from the frailty of the other groups, but this frailty is shared by all individuals within a group (Henderson and Oman, 1999; Hosmer and Lemeshow, 1999; Hougaard, 2000). Individuals of the same group are assumed to be correlated (Cleves *et al.*, 2004; Kleinbaum and Klein, 2005), even though conditional on the frailty they are uncorrelated (Box-Steffensmeier and Zorn, 1999; Gutierrez, 2002; Hougaard, 2000).

The conditional hazard function of the shared frailty model for the individual i in the group j is expressed as (Gutierrez, 2002)

$$h_{ij}\left(t \mid X_{ij}, \alpha_j\right) = \alpha_j h_{ij}\left(t \mid X_{ij}\right) \qquad \textbf{(93),}$$

where α_j is the shared-frailty, *i.e.*, group unobserved heterogeneity. It is also assumed that α_j is a random positive value with mean 1 and variance θ and θ is estimated from the data (Cleves *et al.*, 2004). The frailty variance θ measures the variability of the frailty among groups (Cleves *et al.*, 2004) and the correlation among the individuals of the same group (Kleinbaum and Klein, 2005; Yashin *et al.*, 2001). When $\alpha = 1$ for all groups, then $\theta = 0$, and the frailty model reduces to the ordinary model (without frailty) (Karim, 2008). Moreover, when $\theta = 0$, there is no correlation among the individuals of the same group (Klein and Moeschberger, 1997; Kleinbaum and Klein, 2005). Large values of θ mean that the variability of the frailty among groups is large and the individuals within each group are strongly correlated (Klein and Moeschberger, 1997). Groups with $\alpha_j > 1$ will experience the event earlier than the estimated time from a no-frailty model, and the opposite will happen to groups with $\alpha_j < 1$ (Klein and Moeschberger, 1997).

Once more, the null hypothesis that $\theta = 0$ can be tested in order to evaluate the existence of shared frailty effects (Cleves *et al.*, 2004). Klein and Moeschberger (1997) present a score test for association.

The derivation of the population survival function for shared frailty models is similar to that presented at section 3.6.1.1 for unshared frailty models (Box-Steffensmeier and Jones, 2004). But the interpretation of the unconditional survival and hazard functions in shared frailty models is different from that on unshared frailty models. In shared frailty models, the unconditional functions only represent the population averages if the number of individuals in each group is not correlated with the level of frailty (Kleinbaum and Klein, 2005).

Wienke (2003), Wienke *et al.* (2003), and Karim (2008) mention some handicaps of the shared frailty model. They are:

- The shared frailty model is a common risks' model, and, as such, it is only appropriate for situations where the unobserved covariates are common to all individuals of a given group. In other words, the shared frailty model assumes that the unmeasured risk factors (and, consequently, the frailty effect) are common to all individuals of a given group.;
- It is difficult to distinguish between population heterogeneity and duration dependence;
- As the frailty effect has to be positive, shared frailty models only account for positive association between the individuals of a given group, which may be unreasonable in some situations.

3.6.2. Some frailty distributions

Even though researchers have to postulate in advance the distribution of frailty, there is no theoretical reason to choose any particular distribution (Box-Steffensmeier and Zorn, 1999; Karim, 2008; Wienke *et al.* 2003). This decision is usually based on mathematical and computational convenience (Wienke *et al.*, 2003; Zdravkovic *et al.*, 2004). Nevertheless, no frailty distribution has all desirable properties (Hougaard, 2000).

The estimates of the coefficients of the covariates and of the frailty variance θ vary with the distribution chosen for frailties (Blossfeld *et al.*, 1989; Box-Steffensmeier and Jones, 2004; Heckman and Singer, 1982; Yashin *et al.*, 2001).

Many probability distributions can be used to describe the frailty, provided that they are continuous, only assume positive values and have mean 1 and finite variance θ (Box-Steffensmeier and Jones, 2004; Kleinbaum and Klein, 2005). Nevertheless, it is convenient that the derived survival function is not too complicate to use, because it is a component of the likelihood function (Karim, 2008).

The most widely used distribution of frailty is the gamma distribution (Blossfeld *et al.*, 1989; Box-Steffensmeier and Zorn, 1999; Hosmer and Lemeshow, 1999). This distribution is very popular because its functions are easily derived using a Laplace transformation (Karim, 2008; Wienke *et al.*, 2003), it is simple to interpret and easy to handle mathematically (Glidden, 1998).

A brief description of some possible distributions for frailties is presented below.

3.6.2.1. Gamma distribution

Clayton (1978) proposed the gamma distribution to model random effects. The gamma model is ideal for situations with high late dependence (*i.e.*, if one individual has a long survival time, it is expected that the same happens to the other individuals of the same group) (Hougaard, 2000).

If α follows a gamma distribution with mean 1 and variance θ with the form (Gutierrez, 2002)

$$g(\alpha) = \frac{\alpha^{1/\theta - 1} \exp(-\alpha/\theta)}{\Gamma(1/\theta)\, \theta^{1/\theta}} \quad \textbf{(94).}$$

Then the unconditional survival function of the frailty model is

$$S_\theta(t \mid X) = \left[1 - \theta \ln S(t \mid X)\right]^{-1/\theta} \quad \textbf{(95),}$$

which is equivalent to

$$S_\theta(t \mid X) = L\left(H(t \mid X)\right) = \left[1 + \theta H(t \mid X)\right]^{-1/\theta} \quad \textbf{(96),}$$

where $L\left(H(t \mid X)\right)$ is the Laplace transform of the integrated hazard function. The implied unconditional hazard function is

$$h_\theta(t \mid X) = h(t \mid X)\left[1 - \theta \ln S(t \mid X)\right]^{-1} = h(t \mid X)\left[1 + \theta H(t \mid X)\right]^{-1} \quad \textbf{(97).}$$

The unconditional density function is then

$$f_\theta(t \mid X) = h(t \mid X)\left[1 + \theta H(t \mid X)\right]^{-\frac{1}{\theta} - 1} \quad \textbf{(98).}$$

The gamma frailty model can be estimated by maximum likelihood, using the simple Laplace transform (Wienke *et al.*, 2003). In this way, the frailty term is integrated out, which means that exists an explicit unconditional survival function and the likelihood function can be derived (Wienke *et al.*, 2003). This model can also be estimated using the expectation maximization (EM) algorithm, the penalized partial likelihood method, the penalized likelihood method (Karim, 2008), or bayesian estimation methods (Wienke *et al.*, 2003). Some disadvantages of the EM algorithm are presented in Therneau *et al.* (2000).

The addition of the frailty effect with a gamma distribution converts a PH model into a non-PH model, because the PH assumption is violated, as demonstrated by Kleinbaum and Klein (2005). Klein and Moeschberger (1997) and Henderson and Oman (1999) state that the PH assumption only persists with positive stable frailty models. Considering the unconditional hazard function of two distinct groups of individuals in a gamma frailty model,

$$h_{\theta 1}(t \mid X) = h_1(t \mid X)\left[1 - \theta \ln S_1(t \mid X)\right]^{-1} \qquad \textbf{(99),}$$

$$h_{\theta 2}(t \mid X) = h_2(t \mid X)\left[1 - \theta \ln S_2(t \mid X)\right]^{-1} \qquad \textbf{(100).}$$

Consequently, the hazard ratio is

$$\begin{aligned} HR_{\theta} &= \frac{h_{\theta 2}(t \mid X)}{h_{\theta 1}(t \mid X)} = \frac{h_2(t \mid X)}{h_1(t \mid X)} \times \frac{1 - \theta \ln S_1(t \mid X)}{1 - \theta \ln S_2(t \mid X)} \\ &= \exp[\beta'(X_2 - X_1)] \times \frac{1 - \theta \ln S_1(t \mid X)}{1 - \theta \ln S_2(t \mid X)} \end{aligned} \qquad \textbf{(101),}$$

which is not constant over time; hence, the PH is violated. A similar demonstration can be done with regard to the inverse Gaussian frailty model. In other words, even if the PH assumption holds at an individual level, at a population level this assumption is violated.

On the other hand, Kleinbaum and Klein (2005) demonstrated that if the AFT assumption holds at an individual level, it also holds at a population level in a gamma frailty model. Considering the following unconditional survival function of two distinct groups of individuals in a gamma frailty model

$$S_{\theta 1}(t \mid X) = \left[1 - \theta \ln S_1(t \mid X)\right]^{-1/\theta} \qquad \textbf{(102),}$$

$$S_{\theta 2}(t \mid X) = \left[1 - \theta \ln S_2(t \mid X)\right]^{-1/\theta} \qquad \textbf{(103).}$$

If

$$S_1(t) = S_2(\psi\, t) \qquad \textbf{(104),}$$

then

$$\begin{aligned} S_{\theta 1}(t \mid X) &= \left[1 - \theta \ln S_2(\psi t \mid X)\right]^{-1/\theta} \\ &= S_{\theta 2}(\psi t \mid X) \end{aligned} \qquad \textbf{(105).}$$

3.6.2.2. Inverse Gaussian distribution

The inverse Gaussian distribution used to model random effects was proposed by Hougaard (1986b). The results of the inverse Gaussian distribution are similar to those of the log-normal distribution (Hougaard, 2000; Karim, 2008).

If α follows an inverse-Gaussian distribution with mean 1 and variance θ with the form (Gutierrez, 2002)

$$g(\alpha) = \left(\frac{1}{2\pi\theta\alpha^3}\right)^{1/2} \exp\left\{-\frac{1}{2\theta}\left(\alpha - 2 + \frac{1}{\alpha}\right)\right\} \qquad \textbf{(106),}$$

the unconditional survival function is expressed as

$$S_\theta(t \mid X) = \exp\left\{\frac{1}{\theta}\left\{1 - \left[1 - 2\theta \ln\left(S(t \mid X)\right)\right]^{1/2}\right\}\right\} \qquad \textbf{(107),}$$

the implied unconditional hazard function is

$$h_\theta(t \mid X) = h(t \mid X)\left[1 - 2\theta \ln S(t \mid X)\right]^{-1/2} \qquad \textbf{(108),}$$

and the unconditional density function is given by

$$f_\theta(t \mid X) = \frac{1}{\theta} S_\theta(t \mid X) \times h(t \mid X)\left[1 - 2\theta \ln S(t \mid X)\right]^{-1/2} \qquad \textbf{(109).}$$

3.6.2.3. Log-normal distribution

McGilchrist and Aisbett (1991) proposed the log-normal distribution to model random effects.

Karim (2008) points out that the likelihood function of a survival model with a log-normal frailty cannot be represented, because "the Laplace transformations are theoretically intractable" (p. 9). Wienke *et al.* (2003) also state that an explicit form of the likelihood function does not exist, and in this way, several estimation methods for bivariate frailty models have been proposed. Some possible estimation methods for this kind of models are (i) estimation based on numerical integration in the maximum likelihood method; (ii) some approximations of the maximum likelihood (Hougaard, 2000); (iii) bayesian estimation methods; (iv) restricted maximum likelihood (REML) method; and (v) penalised likelihood method (Hougaard, 2000; Karim, 2008; Wienke *et al.*, 2003).

3.6.2.4. Positive stable distributions

Hougaard (1986a) proposed the positive stable distribution to model random effects. The positive stable distributions are ideal for situations with high early dependence (*i.e.*, if one individual has a short survival time, it is expected that the same happens to the other individuals of the same group) (Hougaard, 2000).

The Laplace transformation cannot be easily used with these models; so, the estimation of these models is much more difficult (Hougaard, 2000). These models can be estimated with the marginal and three-stage methods, both proposed by Hougaard (2000). As these estimation methods are based on approximations, this model is not as popular as the gamma or the log-normal model (Karim, 2008).

3.6.2.5. Power variance function (PVF) distributions

Hougaard (1986b) proposed the power variance model. This family of distributions is a generalization of gamma, inverse Gaussian, and positive stable models (Hougaard, 2000; Karim, 2008). This model can be estimated using the full conditional estimation procedures (Hougaard, 2000).

3.6.3. Advanced frailty models

The frailty models mentioned above assume that the frailty effect is constant over time. But it may happen that the effect of some omitted covariates changes over time (Aalen and Gjessing, 2005). Gjessing *et al.* (2003) propose a flexible generalized frailty model, which considers that frailty is the result of a stochastic process.

Considering the effect of memory about past events on frailty, the following models may result (Aalen and Gjessing, 2005):

- *Standard frailty model* – frailty is determined at the beginning of the follow-up period and do not change over time.
- *Cumulative frailty model – frailty is steadily built up over the individual lifetime and the past frailties are not forgotten.*
- *Moving average frailty model – the past frailties are gradually forgotten and a quasi-stationary process is achieved.*
- *Frailty model with no memory – the past frailties are forgotten and, as such, they do not affect the current frailty.*

Aalen and Gjessing (2005) show that the degree of memory in the frailty process strongly influences the effect of frailty on the hazard function.

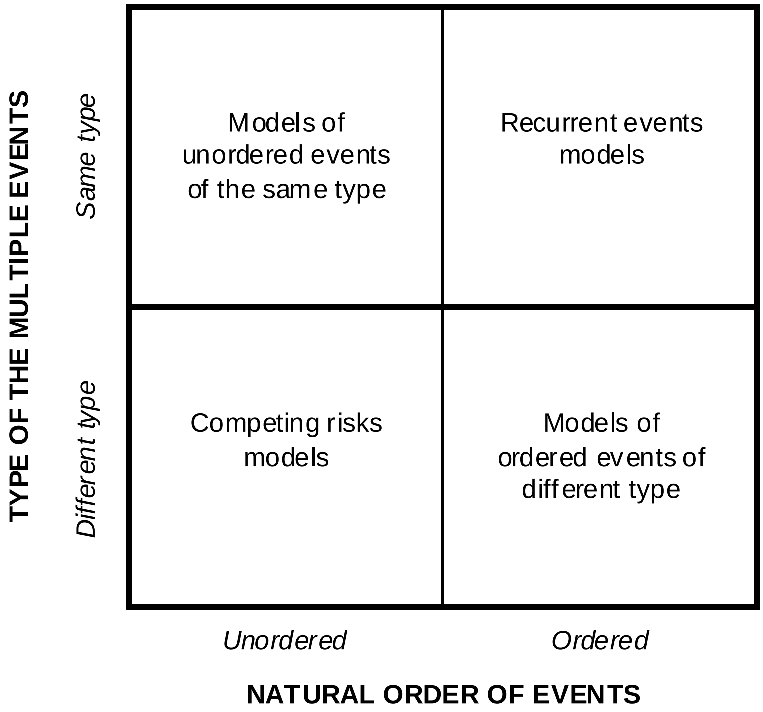
3.7. Multiple events models

The models presented up to this section have considered that there is only one event of interest and that this event can only occur once for each individual (*i.e.*, two-state models or one-way transition models). An important assumption for this kind of models is that survival times are independent. Nevertheless, more complicated situations exist that involve multiple events (*i.e.*, more than one event, of the same or different type, can occur to a given individual). In this situation, the assumption of independent survival times is probably not satisfied (Box-Steffensmeier and Jones, 2004; Cleves, 1999; Kleinbaum and Klein, 2005), because the several survival times for the same individual are probably correlated (the second and posterior events are probably to be affected by the previous events). If this correlation is not considered in the model, the estimates of the coefficients of the covariates are probably biased and the variance estimates could be misleading (Aalen, 1992), because the amount of information about each observation is overstated (Box-Steffensmeier and Zorn, 2002).

Two different approaches have been proposed in the literature to handle dependent survival times, which are (i) shared frailty models, and (ii) robust estimation (Cleves, 1999).

Multiple events can be divided into "ordered" or "unordered" and "same type" or "different type" (Box-Steffensmeier and Jones, 2004; Cleves, 1999). From its combination, four types of models emerge: unordered events of the same type, recurrent events, competing risks, and models of ordered events of different type. Figure 12 shows this classification.

Figure 12 – Types of multiple events models



These types of models are not the aim of the present study, because our database only includes individuals with only a possible event of interest.

3.8. Model diagnostics

The diagnosis of survival models should include an analysis of the functional form of covariates, the validation of the PH assumption, an analysis of the goodness-of-fit of the model, and an identification of outliers and influential observations.

The fit of regression models is analysed based on a comparison of the observed and estimated value of the dependent variable for each individual, that is,

$$Residual_i = y_i - \hat{y}_i \,.$$

Nevertheless, this methodology cannot be directly applied to survival models (Blossfeld *et al.*, 2007, Singer and Willett, 2003), because in survival models the dependent variable is the hazard rate, which is not observable (Blossfeld *et al.*, 2007). One way to extend this idea of regression models to survival models is to choose a quantity to analyse (*e.g.*, survival time, the integrated hazard function, etc.) and develop a strategy that correctly handles censoring (Singer and Willett, 2003). Thus, some different measures have been proposed to analyse the different components of survival models evaluation, and most of them consist in the analysis of different types of residuals.

Even though the model diagnosis methodologies have been mainly developed for the Cox PH model, they can be applied to both semi-parametric and parametric models (Box-Steffensmeier and Jones, 2004).

In continuous survival models, residual analysis is very useful to assess the model adequacy (Box-Steffensmeier and Jones, 2004). An explanation of the most important types of residuals in survival models is presented below.

Cox-Snell residuals

The Cox-Snell residual for the individual i is defined as

$$r_{cs_i} = -\log \hat{S}_i\left(t \mid X\right) = \hat{H}_i\left(t \mid X\right) \qquad \textbf{(110).}$$

Cox-Snell residuals can be used to examine the goodness-of-fit of any parametric and semi-parametric model (Box-Steffensmeier and Jones, 2004; Cleves *et al.*, 2004; Klein and Moeschberger, 1997; Lee and Wang, 2003). These residuals are not symmetrically distributed around zero and they cannot be negative (Collet, 1994).

Schoenfeld residuals

Schoenfeld residuals are the difference between the observed and the expected values of each covariate for each individual at each failure time (Box-Steffensmeier and Jones, 2004; Lee and Wang, 2003). As such, Schoenfeld residuals cannot be computed for censored observations; censored observations only contribute to this computation when they are part of the risk set in the failure time of other individuals (Singer and Willett, 2003).

Specifically, the Schoenfeld residual for the individual i on the covariate k is given by (Hosmer and Lemeshow, 1999)

$$r_{S_i k} = \frac{\partial L}{\partial \beta_k} = \delta_i \left(X_{ik} - \bar{X}_{w_i k} \right) = \delta_i \left(X_{ik} - \frac{\sum_{j \in R(t_i)} X_{jk} \exp(\beta' X)}{\sum_{j \in R(t_i)} \exp(\beta' X)} \right) \qquad \textbf{(111),}$$

where X_{ik} is the value of the covariate k for the individual i, and $R(t_i)$ is the risk set at t_i.

Schoenfeld residuals are not correlated with one another and asymptotically they have mean zero (Lee and Wang, 2003; Singer and Willett, 2003).

The PH assumption can be tested based on Schoenfeld residuals, because the time-dependency of the covariate coefficient can be examined by plotting the Schoenfeld residuals against time (Cleves *et al.*, 2004; Singer and Willett, 2003; Therneau and Grambsch, 2000).

Scaled Schoenfeld residuals

The scaled Schoenfeld residuals are defined as (Hosmer and Lemeshow, 1999)

$$r^*_{S_i} = \left[\hat{V}\left(r_{S_i} \right) \right]^{-1} r_{S_i} \qquad \textbf{(112),}$$

where $\hat{V}\left(r_{S_i}\right)$ is the estimator of the covariance matrix of the vector of Schoenfeld residuals for the individual i, and r_{S_i} are the Schoenfeld residuals for the individual i. The scaled Schoenfeld residuals are useful to evaluate the PH assumption.

Approximated scaled Schoenfeld residuals

The approximated scaled Schoenfeld residuals were proposed by Grambsch and Therneau (1994). They are defined as

$$r_{S_i}^{**} = m\hat{V}\left(\hat{\beta}\right) r_{S_i} \qquad \textbf{(113),}$$

where m is the observed number of uncensored survival times, and $\hat{V}\left(\hat{\beta}\right)$ is the estimator of the covariance matrix of the estimated coefficients. These residuals are useful to assess the PH assumption.

Martingale residuals

Martingale residuals are the difference between the observed and the expected number of events for an individual based on the estimated model (Therneau *et al.*, 1990), and are expressed as (Therneau and Grambsch, 2000)

$$M_i = N_i\left(t\right) - \hat{H}_i\left(t \mid X\right) \qquad \textbf{(114),}$$

where $N_i\left(t\right)$ is the observed number of events for the individual i at time t, and $\hat{H}_i\left(t \mid X\right)$ is the expected number of events for the individual i at time t, based on the estimated model.

When the event of interest can only occur once, Equation 114 is equivalent to (Collet, 1994)

$$M_i = \delta_i - \hat{H}_i(t \mid X) = \delta_i - r_{cs_i} \qquad (115).$$

Martingale residuals are the error component in the counting process and, as such, it is proved that $E(M_i) = 0$, $\mathrm{cov}(M_i, M_j) = 0, \quad i \neq j$, and $\sum M_i = 0$ in large samples (Collet, 1994; Therneau and Grambsch, 2000; Therneau *et al.*, 1990):

Martingale residuals vary in the interval $]-\infty, N_i[$; so, they are highly skewed and not symmetric around zero as the usual residuals in linear models (Box-Steffensmeier and Jones, 2004; Collet, 1994; Therneau and Grambsch, 2000).

Negative values of martingale residuals arise when the individual experienced fewer events than expected or when the individual experienced the events later than expected (Box-Steffensmeier and Jones, 2004; Lee and Wang, 2003; Singer and Willet, 2003), which means that the model underpredicts (Singer and Willet, 2003). On the other hand, martingale residuals are positive when the event occurred earlier than expected (Box-Steffensmeier and Jones, 2004; Lee and Wang, 2003; Singer and Willet, 2003). In these circumstances, the model overpredicts (Singer and Willet, 2003). As can be seen in Equation 115, martingale residuals of censored observations are always negative (Collet, 1994; Tableman and Kim, 2004).

Martingale residuals are a useful tool to examine the functional form of the covariates of the model (Box-Steffensmeier and Jones, 2004; Cleves *et al.*, 2004; Collet, 1994; Hosmer and Lemeshow, 1999; Klein and Moeschberger, 1997; Singer and Willett, 2003; Therneau and Grambsch, 2000).

Deviance residuals

Deviance residuals were introduced by Therneau *et al.* (1990). Deviance residuals are a mere transformation of the martingale residuals in order to obtain a more normal distribution when the appropriate model is specified (Collet, 1994; Klein and Moeschberger, 1997; Therneau *et al.*, 1990), even though the sum of deviance residuals is not necessarily zero (Collet, 1994). When less than about a quarter of the total observations are censored, deviance residuals have a distribution very close to the normal distribution (Therneau *et al.*, 1990). But when there is heavy censoring, a cloud of points with close to zero residuals will distort the normal distribution (Therneau *et al.*, 1990). According to Singer and Willet (2003), heavy censoring is presented when more than approximately 40% of the total observations are censored.

The interpretation of the deviance residuals is similar to that of martingale residuals (Singer and Willet, 2003). The deviance residual is zero only when the martingale residual is zero (Therneau *et al.*, 1990). The deviance residuals of censored observations are always negative.

When covariates are time-invariant and the event of interest only occurs once, deviance residuals have the following form (Therneau *et al.*, 1990)

$$D_i = sign\left(M_i\right)\left\{-2\left[\,M_i + \delta\log\left(\,\delta - M_i\right)\,\right]\right\}^{1/2} \qquad \textbf{(116),}$$

where $sign(.)$ is the sign function.

The explanation of the survival model diagnostics is presented in the Chapter 4.

4. DATA AND RESULTS

4.1. Introduction

As mentioned before, the aim of this study is to analyse the partial churn of residential customers in the fixed-telecommunications industry. Specifically, this is a longitudinal study of the probability of fixed-telephone and ADSL contracts cancellation at time t, given that the contracts last until t and given some covariates.

Section 4.2 describes the data in which the empirical study is based. After that, models for the partial customer churn of fixed-telephone and ADSL contracts are developed for customers who have (or had) both types of contracts in sections 4.3 and 4.4, respectively. Two-state models are used as the event of interest occurs only once to each individual. Lastly, the effect of customer satisfaction on the cancellation of fixed-telephone and ADSL contracts are tested in sections 4.5 and 4.6, respectively.

4.2. Data

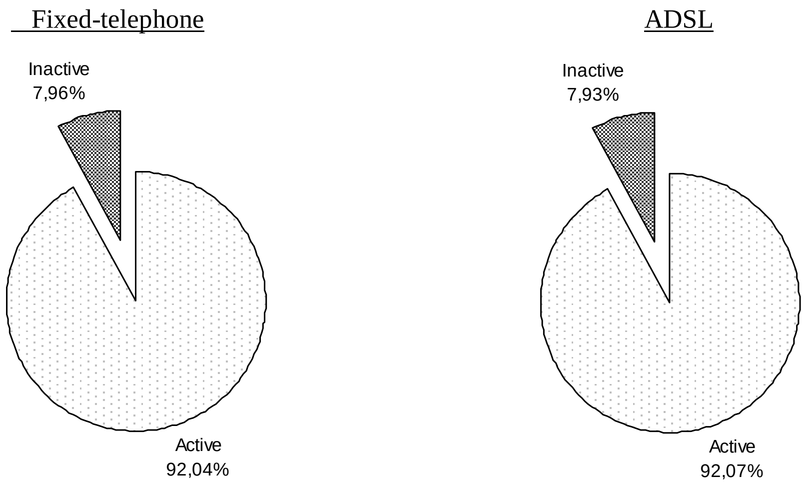
Data were obtained from a Portuguese fixed telecommunications firm which offers fixed telephone, ADSL, pay-TV and home-video. The present study focus on the analysis of customers who have both fixed telephone and ADSL contracts. The time window of analysis is from March 2003 until November 2008. Only the geographic area of Portugal Continental is studied. Customers are observed from the time they contract a service from the firm until the time they cancel all contracts with the firm or until the end of observation period (November 2008).

Each contract has an initial subscription period. The subscription period may vary across types of services and even across customers. Customers cannot cancel the contract within this initial subscription period. If customers decide to cancel the contract within this period, they have to pay the remaining amount of the period. After this initial subscription period, customers are allowed to cancel the contract at any time, without any penalty.

The study is developed by using a large-scale database with residential customers who have active contracts with the firm during at least one month between 1st December 2007 and 30th November 2008. This condition was imposed because the variables about usage and revenues were only available for this period of time.

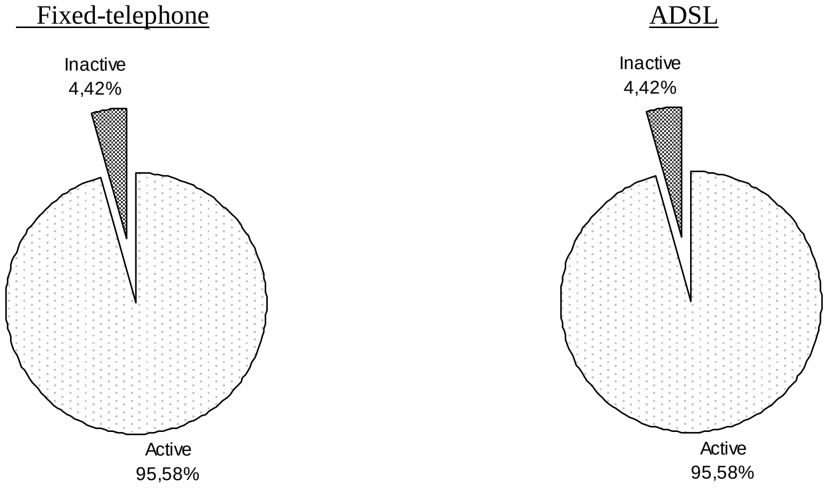
The first part of the empirical study will separately estimate the hazard function of ADSL contracts and fixed-telephone contracts for residential customers who have both contract types. This includes almost 80.000 customers. Figure 13 shows the percentage of active and inactive contracts of ADSL and fixed-telephone services, respectively.

Figure 13 - Active and inactive contracts in fixed-telephone and ADSL services (population)



The second part of the empirical study intends to examine the effect of customer satisfaction on the customer hazard function of fixed-telephone and ADSL contracts for residential customers. This analysis was based on a database that includes a random sample of about 700 residential customers who completed a questionnaire about customer satisfaction. The percentage of active and inactive contracts of ADSL and fixed-telephone services included in this random sample is shown in Figure 14.

Figure 14 - Active and inactive contracts in fixed-telephone and ADSL services (sample)



As can be seen in the Figures presented above, the percentage of inactive customers is very low, which means that heavy censoring is presented in this study. This situation is consistent with several studies about customer churn prediction (*e.g.*, Ahn *et al.*, 2006 ; Bin *et al.*, 2007; Bonfrer *et al.*, 2007; Ferreira *et al.*, 2004; Hung *et al.*, 2006; and Wei and Chiu 2002).

The database provided by the firm contains a large number of covariates, which include information about the contract, customer demographics, payment history, customer historical information about global revenues, average revenues from the fixed telephone service and from the ADSL service from December 2007 until November 2008, and average usage of fixed telephone and ADSL from December 2007 until November 2008.

As presented in the literature, other variables might be important for estimating the hazard functions (for instance, the subscription period of each contract, promotions, acquisition cost, contact details to and from the customer, complaints, customer satisfaction, other demographic data such as age, education, number of people in the household, etc). Nevertheless, it is believed that accurate hazard models can be estimated with the available data.

The database provided by the firm was modified, by transforming the nominal variables into multiple binary variables and some binary variables were created based on some existent variables. Table 6 presents a description of the resulting database.

Table 6 – Description of the database

		Variables	Description
ID	1	Cust_id	Customer number
	2	Contract_id	Contract number
Contract	3	Contract_lifetime	Duration of the contract (in days)
	4	Contract_activation_date	Contract activation date
	5	Contract_desactivation_date	Contract desactivation date
	6	Contract_status	Contract status (0 – active; 1 – inactive)
	7	Product	Type of product (0 – fixed-telephone; 1 – ADSL)
	8	Portability	Portability of the telephone number (0 – no; 1 – yes)
	9	Payment_method	Payment method (0 – direct debit; 1 – other)
	10	Flat_plan_teleph_1	Have the customer contracted the flat plan fixed-telephone type 1 (fixed telephone)? (0 – no; 1 – yes)
	11	Flat_plan_teleph_2	Have the customer contracted the flat plan fixed-telephone type 2 (fixed telephone)? (0 – no; 1 – yes)
	12	Flat_plan_teleph_3	Have the customer contracted the flat plan fixed-telephone type 3 (fixed telephone)? (0 – no; 1 – yes)
	13	Flat_plan_ADSL_1	Have the customer contracted the flat plan ADSL type 3 (fixed telephone)? (0 – no; 1 – yes)
	14	Equipment_renting	Have the customer rented any equipment? (0 – no; 1 – yes)
Customer demographics	15	Gender	Gender of the customer (0 – female; 1 – male)
	16	Aveiro	Does the customer live in province “Aveiro”? (0 – no; 1 – yes)
	17	Beja	Does the customer live in province “Beja”? (0 – no; 1 – yes)
	18	Braga	Does the customer live in province “Braga”? (0 – no; 1 – yes)
	19	Bragança	Does the customer live in province “Bragança”? (0 – no; 1 – yes)
	20	Castelo Branco	Does the customer live in province “Castelo Branco”? (0 – no; 1 – yes)
	21	Coimbra	Does the customer live in province “Coimbra”? (0 – no; 1 – yes)
	22	Évora	Does the customer live in province “Évora”? (0 – no; 1 – yes)
	23	Faro	Does the customer live in province “Faro”? (0 – no; 1 – yes)

		Variables	Description
	24	Guarda	Does the customer live in province “Guarda”? (0 – no; 1 – yes)
	25	Leiria	Does the customer live in province “Leiria”? (0 – no; 1 – yes)
	26	Lisboa	Does the customer live in province “Lisboa”? (0 – no; 1 – yes)
	27	Portalegre	Does the customer live in province “Portalegre”? (0 – no; 1 – yes)
	28	Porto	Does the customer live in province “Porto”? (0 – no; 1 – yes)
	29	Santarém	Does the customer live in province “Santarém”? (0 – no; 1 – yes)
	30	Setúbal	Does the customer live in province “Setúbal”? (0 – no; 1 – yes)
	31	Viana do Castelo	Does the customer live in province “Viana do Castelo”? (0 – no; 1 – yes)
	32	Vila Real	Does the customer live in province “Vila Real”? (0 – no; 1 – yes)
	33	Viseu	Does the customer live in province “Viseu”? (0 – no; 1 – yes)
Payment history	34	N_total_dunning	Total number of overdue bills since the beginning of the contract
	35	Current_debts	Value of current debts (in euros)
Global revenues	36	Mean_overall_revenues	Monthly mean of the total revenues from the customer since the beginning of the contract (in euros)
	37	Mean_revenues	Monthly average revenues from the customer between December 2007 and November 2008 (in euros)
Fixed telephone revenues	38	Mean_int_out_value	Monthly average value of international calls (outside the pack) between December 2007 and November 2008 (in euros)
	39	Mean_loc_out_value	Monthly average value of local calls (outside the pack) between December 2007 and November 2008 (in euros)
	40	Mean_nat_out_value	Monthly average value of national calls (outside the pack) between December 2007 and November 2008 (in euros)
	41	Mean_mobile_value	Monthly average value of calls to mobile phones between December 2007 and November 2008 (in euros)
	42	Mean_other_value	Monthly average value of other kind of calls between December 2007 and November 2008 (in euros)
	43	Mean_loc_peak_value	Monthly average value of local calls (peak time) between December 2007 and November 2008 (in euros)
	44	Mean_loc_off_peak_value	Monthly average value of local calls (off-peak time) between December 2007 and November 2008 (in euros)
	45	Mean_nat_peak_value	Monthly average value of national calls (peak time) between December 2007 and November 2008 (in euros)

		Variables	Description
	46	Mean_nat_off_peak_value	Monthly average value of national calls (off-peak time) between December 2007 and November 2008 (in euros)
	47	Mean_value_calls_offpeak	Monthly average value of calls (off-peak time) between December 2007 and November 2008 (in euros)
	48	Mean_value_calls_peak	Monthly average value of calls (peak time) between December 2007 and November 2008 (in euros)
	49	Mean_calls_revenues	Monthly average revenues from the fixed-telephone service between December 2007 and November 2008 (in euros)
	50	Mean_revenues_flat_plan_teleph_1	Monthly average revenues from the flat plan type 1 (fixed telephone) between December 2007 and November 2008 (in euros)
	51	Mean_revenues_flat_plan_teleph_2	Monthly average revenues from the flat plan type 2 (fixed telephone) between December 2007 and November 2008 (in euros)
	52	Mean_revenues_flat_plan_teleph_3	Monthly average revenues from the flat plan type 3 (fixed telephone) between December 2007 and November 2008 (in euros)
	53	Mean_revenues_equipm_renting	Monthly average revenues from the equipment renting between December 2007 and November 2008 (in euros)
	54	Mean_revenues_flat_plan_ADSL_1	Monthly average revenues from the flat plan type 1 (ADSL) between December 2007 and November 2008 (in euros)
ADSL revenues	55	Mean_ADSL_revenues	Monthly average total revenues from ADSL service between December 2007 and November 2008 (in euros)
	56	Mean_value_additional_traffic	Monthly average value of additional internet traffic between December 2007 and November 2008 (in euros)
Fixed- telephone usage	57	Mean_int_out_duration	Monthly average duration of international calls (outside the pack) between December 2007 and November 2008 (in minutes)
	58	Mean_int_in_duration	Monthly average duration of international calls (inside the pack) between December 2007 and November 2008 (in minutes)
	59	Mean_int_out_quantity	Monthly average number of international calls (outside the pack) between December 2007 and November 2008
	60	Mean_int_in_quantity	Monthly average number of international calls (inside the pack) between December 2007 and November 2008
	61	Mean_loc_out_duration	Monthly average duration of local calls (outside the pack) between December 2007 and November 2008 (in minutes)
	62	Mean_loc_in_duration	Monthly average duration of local calls (inside the pack) between December 2007 and November 2008 (in minutes)
	63	Mean_loc_out_quantity	Monthly average number of local calls (outside the pack) between December 2007 and November 2008
	64	Mean_loc_in_quantity	Monthly average number of local calls (inside the pack) between December 2007 and November 2008

	Variables	Description
65	Mean_nat_out_duration	Monthly average duration of national calls (outside the pack) between December 2007 and November 2008 (in minutes)
66	Mean_nat_in_duration	Monthly average duration of national calls (inside the pack) between December 2007 and November 2008 (in minutes)
67	Mean_nat_out_quantity	Monthly average number of national calls (outside the pack) between December 2007 and November 2008
68	Mean_nat_in_quantity	Monthly average number of national calls (inside the pack) between December 2007 and November 2008
69	Mean_mobile_duration	Monthly average duration of calls to mobile phones between December 2007 and November 2008 (in minutes)
70	Mean_mobile_quantity	Monthly average number of calls to mobile phones between December 2007 and November 2008
71	Mean_other_duration	Monthly average duration of other kind of calls between December 2007 and November 2008 (in minutes)
72	Mean_other_quantity	Monthly average number of other kind of calls between December 2007 and November 2008
73	Mean_loc_peak_duration	Monthly average duration of local calls (peak time) between December 2007 and November 2008 (in minutes)
74	Mean_loc_off_peak_duration	Monthly average duration of local calls (off-peak time) between December 2007 and November 2008 (in minutes)
75	Mean_loc_peak_quantity	Monthly average number of local calls (peak time) between December 2007 and November 2008
76	Mean_loc_off_peak_quantity	Monthly average number of local calls (off-peak time) between December 2007 and November 2008
77	Mean_nat_peak_duration	Monthly average duration of national calls (peak time) between December 2007 and November 2008 (in minutes)
78	Mean_nat_off_peak_duration	Monthly average duration of national calls (off-peak time) between December 2007 and November 2008 (in minutes)
79	Mean_nat_peak_quantity	Monthly average number of national calls (peak time) between December 2007 and November 2008
80	Mean_nat_off_peak_quantity	Monthly average number of national calls (off-peak time) between December 2007 and November 2008
81	Mean_duration_calls_offpeak	Monthly average duration of off-peak calls between December 2007 and November 2008 (in minutes)
82	Mean_duration_calls _peak	Monthly average duration of peak calls between December 2007 and November 2008 (in minutes)
83	Mean_duration_calls_in	Monthly average duration of calls (inside the pack) between December 2007 and November 2008 (in minutes)
84	Mean_duration_calls _out	Monthly average duration of calls (outside the pack) between December 2007 and November 2008 (in minutes)
85	Mean_duration_calls_total	Monthly average duration of calls (total) between December 2007 and November 2008 (in minutes)
86	Mean_quantity_calls_in	Monthly average number of calls (inside the pack) between December 2007 and November 2008

		Variables	Description
	87	Mean_quantity_calls_out	Monthly average number of calls (outside the pack) between December 2007 and November 2008
	88	Mean_quantity_calls_total	Monthly average umber of calls (total) between December 2007 and November 2008
	89	Mean_quantity_calls_offpeak	Monthly average number of calls (off-peak time) between December 2007 and November 2008
	90	Mean_quantity_calls_peak	Monthly average number of calls (peak time) between December 2007 and November 2008
ADSL usage	91	Mean_internet_traffic	Monthly average internet traffic between December 2007 and November 2008 (in gigabytes)
	92	Mean_additional_traffic	Monthly average additional internet traffic between December 2007 and November 2008 (in gigabytes)

4.3. Partial customer churn: The fixed-telephone contracts

4.3.1. Selection of covariates

Considering the large number of available variables about customers and that some of them are probably correlated, the correlation matrix was computed in order to decide which covariates to include in the models. Table 7 presents the selected covariates to be used in the hazard model of the fixed-telephone service.

Table 7 - Selected covariates to the hazard model of fixed-telephone contracts

		Covariates
Contract	1	Portability
	2	Payment_method
	3	Flat_plan_teleph_1
	4	Flat_plan_teleph_2
	5	Flat_plan_teleph_3
	6	Equipment_renting
Customer demographics	7	Gender
	8	Beja
	9	Braga
	10	Castelo Branco
	11	Coimbra
	12	Évora
	13	Faro
	14	Guarda
	15	Leiria
	16	Lisboa
	17	Portalegre
	18	Porto
	19	Santarém
	20	Setúbal
	21	Viana do Castelo
	22	Vila Real
	23	Viseu
Payment history	24	N_total_dunning
	25	Current_debts
Global revenues	26	Mean_overall_revenues
Fixed telephone revenues	27	Mean_int_out_value
	28	Mean_loc_out_value
	29	Mean_nat_out_value
	30	Mean_mobile_value
	31	Mean_other_value
Fixed-telephone usage	32	Mean_int_in_duration
	33	Mean_loc_in_duration
	34	Mean_nat_in_duration
	35	Mean_other_duration
	36	Mean_quantity_calls_out

4.3.2. Analysis of the functional form of covariates

The analysis of the functional form of the covariates is of great importance, because when the functional form of covariates is misspecified, the estimated coefficients of the covariates are biased (Keele, 2008). As mentioned in section 3.8, martingale residuals can be used to analyse the functional form of covariates. Even though several methods based on the residuals have been proposed to analyse the functional form of covariates (*e.g.*, Box-Steffensmeier and Jones, 2004; Grambsch *et al.*, 1995; Therneau *et al.*, 1990), Therneau and Grambsch (2000) point out that the simplest method to examine the functional form of a given covariate is to plot the smoothed curve of the martingale residuals from a null model against the values of that covariate. The form of the smoothed curve indicates the functional form of that covariate. Nevertheless, they point out that this method may fail when covariates are correlated. These authors suggest smoothing the martingale residuals with the locally weighted scatterplot smoothing (lowess) method. Thus, in this study, the functional form of continuous covariates is analysed as proposed by Therneau and Grambsch (2000). These graphs are shown in Appendix B. From the analysis of these graphs, it can be concluded that the plots are approximately linear, and no known transformation id required.

4.3.3. Testing the PH assumption

The Cox model has been largely applied in situations where the PH assumption is far from being satisfied (Schemper, 1992). Nevertheless, the violation of the PH assumption originates the following consequences for the results of the model:

- As regards to the covariates that do not satisfy the PH assumption, the power of the corresponding statistical tests decreases (Lagakos and Schoenfeld, 1984);
- As regards to the covariates that satisfy the PH assumption, the power of the corresponding statistical tests also decreases due to a low fit of the model (Schemper, 1992);
- The estimates of the coefficients of the covariates are biased. Thus, the estimates of the coefficients of the covariates with hazard ratios increasing over times are overestimated, and the estimates of the coefficients of the covariates with converging hazard ratios (probably the most common violation) are underestimated (Schemper, 1992).

As such, when the PH assumption is violated, the model is invalid (Hess, 1995). A correct interpretation of the coefficients of a PH model can only be made when this assumption holds (Parmar and Machin, 1995).

Statistical failure of the PH assumption may be due to the existence of some other problems in the model specification, such as the functional form of the covariates (Keele, 2008; Therneau and Grambsch, 2000), and, as such, Keele (2008) suggests the correction of these misspecifications before the evaluation of the PH assumption. In fact, in presence of these misspecifications, the statistical tests for the evaluation of the PH assumption provide evidence about non-PH, when the model is PH (Keele, 2008). Keele (2008) shows that the correction of the functional form of covariates modifies the diagnosis about the PH assumption.

In some circumstances, the statistical failure of the PH assumption is not a big problem, particularly in large samples, when the graphical approach shows that the model is almost PH, and, for instance, the reason of failure is the presence of outliers (Therneau and Grambsch, 2000).

Nevertheless, when non-PH is effective, the problem cannot be ignored and some strategies should be adopted in order to overtake it, as proposed by Box-Steffensmeier and Zorn (2001), Collet (1994), Schemper (1992), and Therneau and Grambsch (2000):

- Fit a stratified Cox model (the stratification variable is the covariate that does not hold the PH assumption) rather then an ordinary Cox model;
- Partition of the survival time axis and fit separate models for each part (piecewise model);
- Inclusion of interaction of time-invariant covariates and some function of time;
- Use AFT or additive hazards model.[5]

Ng'Andu (1997) also proposes some strategies of modelling in presence of non-PH.

---

[5] The Aalen additive hazard model assume that covariates have an additive effect on the hazard function and this effect may not be constant (in absolute value and sign) over time (Bradburn *et al.*, 2003). The hazard function of the additive model is given by $h(t \mid X) = h_0(t) + \beta'(t)X$. Even though this approach seems to be more flexible, its interpretation is complicate (Bradburn *et al.*, 2003). Moreover, this approach is not largely implemented in the commercial statistical software (Bradburn *et al.*, 2003).

Some methods are available to test the PH assumption. Box-Steffensmeier and Zorn (2001) mention three types of methods to test the PH assumption, namely (i) piecewise regressions to detect changes in parameter values; (ii) residual based tests (graphical and statistical); and (iii) tests of the coefficients of the interaction of covariates with time. Additionally, Lee and Wang (2003) suggest a comparison of the fit of the PH and non-PH models. Ata and Sozer (2007) point out that there is not the best approach to examine the PH assumption. Ng'Andu (1997) compared five statistical tests whose objective is to test the PH assumption, with and without censoring, and he found that the interaction of time-invariant covariates and time and the Grambsch and Therneau (1994) test are equally powerful in detecting non-PH. Furthermore, he concluded that statistical tests based on the partition of survival time have less power than the statistical tests that do no request the partition of survival time. Lastly, he provided evidence that the power of the statistical tests depends on the magnitude and type of divergence from the PH. Box-Steffensmeier and Jones (2004) point out that the piecewise regression is the best method to test for the PH assumption in parametric models, even though there are better methods to test the PH assumption in Cox models.

Thus, in this study, the PH assumption is tested by using five different methods, which are: piecewise regression, statistical tests and graphical approaches, testing the coefficients of the interaction of time-invariant covariates with a function of time, and lastly, comparing the fitting of PH and non-PH models.

4.3.3.1. Piecewise regression

In order to test the PH assumption based on piecewise regression, the dataset must be divided into at least two groups; one with individuals that survive less than a given value (*e.g.*, the median survival time) and other group with individuals that survival more than that value. Then, separate survival models are fitted for each group of observations (Box-Steffensmeier and Jones, 2004; Hess, 1995). If the estimated coefficients of the covariates are consistent across the two models, the PH assumption is satisfied (Box-Steffensmeier and Zorn, 2001; Box-Steffensmeier and Jones, 2004); otherwise, the PH assumption is violated.

The number and location of the breakpoints have been widely discussed in the literature (Box-Steffensmeier and Zorn, 2001). It is recommended that each interval of time has similar number of events (*i.e.*, the quantiles of survival time may be used to create groups of individuals) and that no interval has few events (Box-Steffensmeier and Zorn, 2001).

As such, our database was divided into two groups. The first group includes the contracts whose lifetime is until the median lifetime (inclusive) and the second group includes the remaining contracts. The median lifetime of fixed-telephone contracts is 783 days. As suggested by Hess (1995) and Box-Steffensmeier and Jones (2004), Cox models are estimated for each group and the coefficients of the covariates are compared. The results of the models are presented in Table 8.

From the analysis of the Table 8, it can be concluded that the estimated coefficients of many covariates differ across groups and as regards to the majority of covariates that are significant in both models, their coefficients are not consistent across the two groups and there are even situations where the signs of the estimated coefficients are the opposite (*payment_method* and *flat_plan_teleph_3*).

As such, it can be said that there is empirical evidence that the effect of some covariates on the cancellation of telephone-fixed contracts is not constant over time, which means that the PH assumption is not satisfied.

Table 8– Estimates of the piecewise models of fixed-telephone contracts

	Group 1		Group 2	
	β	***p*-value**	β	***p*-value**
N_total_dunning	1.733	0.000**	3.040	0.000**
Mean_overall_revenues	0.017	0.000**	0.003	0.143
Current_debts	-0.027	0.000**	-0.083	0.000**
Mean_int_out_value	-0.011	0.000**	0.005	0.464
Mean_int_in_duration	-0.001	0.000**	-0.001	0.250
Mean_loc_out_value	-0.007	0.681	0.001	0.960
Mean_loc_in_duration	0.000	0.886	0.000	0.786
Mean_nat_out_value	0.091	0.006**	-0.051	0.141
Mean_nat_in_duration	0.000	0.019*	0.000	0.126
Mean_mobile_value	0.006	0.118	0.005	0.331
Mean_other_value	0.088	0.000**	0.080	0.000**
Mean_other_duration	0.001	0.000**	0.000	0.418
Mean_quantity_calls_out	-0.001	0.355	0.004	0.048*
Portability	0.406	0.000**	0.300	0.000**
Payment_method	0.389	0.000**	-0.509	0.000**
Flat_plan_teleph_1	-0.524	0.000**	0.170	0.125
Flat_plan_teleph_2	-0.632	0.000**	-0.457	0.000**
Flat_plan_teleph_3	-0.398	0.000**	0.620	0.000**
Equipment_renting	0.187	0.074	0.081	0.735
Beja	-0.440	0.145	0.219	0.612
Braga	-0.149	0.221	-0.412	0.008**
Castelo Branco	-0.385	0.172	-0.535	0.214
Coimbra	-0.211	0.085	0.063	0.736
Évora	-0.400	0.204	-0.491	0.499
Faro	-0.121	0.331	0.025	0.899
Guarda	-0.390	0.199	0.125	0.809
Leiria	-0.381	0.006**	-0.102	0.615
Lisboa	-0.329	0.000**	-0.301	0.032*
Portalegre	-0.210	0.288	-0.698	0.333
Porto	-0.094	0.313	-0.304	0.032*
Santarém	-0.446	0.000**	0.176	0.376
Setúbal	-0.139	0.194	0.102	0.519
Viana do Castelo	-0.510	0.002**	-0.315	0.207
Vila Real	-0.408	0.068	0.043	0.905
Viseu	-0.206	0.363	-0.782	0.277
Gender	-0.005	0.898	-0.004	0.925

** significant at the 1% level; * significant at the 5% level

4.3.3.2. Statistical tests based on residuals

Several statistical tests have been proposed in the literature to test the PH assumption. Grambsch and Therneau (1994) proposed a global test of the PH assumption. It is a global test because it tests the model as a whole and not each covariate in separate. Under the null hypothesis that hazards are proportional, it is expected that the correlation between the Schoenfeld residuals and survival time is zero (Cleves *et al.*, 2004). Thus, if one concludes that the PH is violated, the covariate(s) which has(ve) problems is unknown; so, Box-Steffensmeier and Jones (2004) argue that each covariate must be examined in separate.

Therneau and Grambsch (1994) propose a Rao efficient score test to examine the PH assumption for each covariate in separate, which is based on the following equation (Therneau and Grambsch, 2000):

$$E\left(r^*_{S_{ik}}\right) + \hat{\beta}_k \approx \beta_k\left(t_i\right) \quad \textbf{(117),}$$

where $r^*_{S_{ik}}$ is the scaled Schoenfeld residual of the individual i for the covariate k, and $\hat{\beta}_k$ is the estimated coefficient of the covariate k from an ordinary Cox model.

Many other tests have been proposed in the literature for the PH assumption, as presented in Therneau and Grambsch (2000). The main difference between them is the function of time that is used (*e.g.*, t, $\log t$, or even a piecewise function) (Therneau and Grambsch, 2000). These different tests may lead to different conclusions about the PH of a given covariate (Kleinbaum and Klein, 2005).

In the present study, the PH assumption is tested for each covariate with the Rao efficient score test of Therneau and Grambsch and (1994). The PH assumption of the global model was tested using the Grambsch and Therneau (1994) test. Table 9 shows the results of both tests.

Table 9– Statistical tests of the PH assumption of fixed-telephone contracts

	Rho	Chi2	Df	*p*-value
N_total_dunning	0.290	1352.90	1	0.000**
Mean_overall_revenues	-0.021	5.66	1	0.017*
Current_debts	-0.266	3801.94	1	0.000**
Mean_int_out_value	-0.033	6.65	1	0.010**
Mean_int_in_duration	0.009	0.83	1	0.362
Mean_loc_out_value	0.076	48.52	1	0.000**
Mean_loc_in_duration	-0.022	2.60	1	0.107
Mean_nat_out_value	0.007	0.44	1	0.508
Mean_nat_in_duration	-0.001	0.01	1	0.943
Mean_mobile_value	-0.018	1.59	1	0.207
Mean_other_value	-0.039	5.68	1	0.017*
Mean_other_duration	-0.011	0.57	1	0.451
Mean_quantity_calls_out	0.007	0.35	1	0.555
Portability	0.196	236.61	1	0.000**
Payment_method	-0.196	374.22	1	0.000**
Flat_plan_teleph_1	0.066	27.58	1	0.000**
Flat_plan_teleph_2	0.036	8.42	1	0.004**
Flat_plan_teleph_3	0.124	99.97	1	0.000**
Equipment_renting	-0.024	3.56	1	0.059
Beja	0.028	5.15	1	0.023*
Braga	0.036	9.10	1	0.003**
Castelo Branco	0.016	1.54	1	0.215
Coimbra	0.043	12.67	1	0.000**
Évora	0.022	3.09	1	0.079
Faro	0.047	15.06	1	0.000**
Guarda	0.024	3.71	1	0.054
Leiria	0.052	17.80	1	0.000**
Lisboa	0.056	21.77	1	0.000**
Portalegre	0.021	2.90	1	0.089
Porto	0.023	3.79	1	0.051
Santarém	0.056	21.82	1	0.000**
Setúbal	0.051	18.46	1	0.000**
Viana do Castelo	0.022	3.23	1	0.072
Vila Real	0.002	0.02	1	0.878
Viseu	0.011	0.73	1	0.392
Gender	0.030	5.64	1	0.018*
Global test		4353.76	36	0.000**

** significant at the 1% level; * significant at the 5% level

These statistical tests provide evidence that the PH assumption does not hold for about 55% of the covariates, and, consequently, on the whole, the model is not PH.

4.3.3.3. Graphical approaches based on residuals

Kleinbaum and Klein (2005) point out that even though statistical tests provide a more objective decision about the PH assumption, graphical approaches allow for detecting some particular deviations from the PH assumption; so they suggest the use of both methods. Also Therneau and Grambsch (2000) emphasise that graphical approaches and statistical tests must be complementary in examination of the PH assumption, because plots allow us to have an idea about the reason and magnitude of failure of this assumption.

There are several graphical methods proposed in the literature to test the PH assumption. For instance, Hess (1995) presents a review of literature of eight distinct graphical methods.

Two graphical approaches are used to verify the PH assumption in this study. Firstly, the Schoenfeld residuals of each covariate (discrete and continuous) were plotted against the survival time. Appendix C shows these graphs. Even though the graphical analysis may be not very objective (Hosmer and Lemeshow, 1999), some curves show a clear trend over time, and, thus, the effect of the respective covariates changes over time, which means that the PH assumption may probably not hold (Box-Steffensmeier and Jones, 2004; Cleves *et al.*, 2004; Hess, 1995; Hosmer and Lemeshow, 1999; Singer and Willett, 2003). These covariates are *Current_debts*, *Portability*, *Payment_method*, *Flat_plan_teleph_1, and Flat_plan_teleph_3*.

Box-Steffensmeier and Jones (2004) suggest the use of statistical tests for the covariates which plot does not show an obvious slope.

The second graphical approach used in this study to verify the PH assumption is the plot of $-\ln\left\{-\ln\left[\hat{S}(t)\right]\right\}$ against $\ln(t)$ for each level of a discrete covariate (where $\hat{S}(t)$ is the Kaplan-Meier estimate of the survival function) (Cleves *et al.*, 2004). PH assumption is satisfied if parallel curves are obtained (Blossfeld *et al.*, 2007; Collet, 1994; Therneau and Grambsch, 2000). Kleinbaum and Klein (2005) demonstrate this property. Therneau and Grambsch (2000) highlight that this method can only be used when covariates are discrete and with few levels. These graphs are presented in Appendix D. As can be seen in the graph of the provinces, nothing can be concluded due to the multiple categories of this covariate. As regards to the remaing discrete covariates, it may be concluded that the PH assumption probably fails in the variables *Portability, Payment_method, Flat_plan_teleph_1, Flat_plan_teleph_2, and Flat_plan_teleph_3,* because the curves are not parallel.

It should be emphasised that the evaluation of the PH assumption by graphical methods has a great drawback, because the decision about "how parallel is parallel" is very subjective and sometimes even difficult to visualise (Ata and Sozer, 2007; Hosmer and Lemeshow, 1999; Kleinbaum and Klein, 2005).

4.3.3.4. Testing the coefficients of the interaction of time-invariant covariates and the functions of time t and $\ln(t)$

Based on the fact that when the effect of one covariate varies with time, the PH assumption is violated, Cox (1972) propose the inclusion in the model of interaction terms between time-invariant covariates and some function of time in order to test the PH assumption. If the estimated coefficient of any of these interactions is statistically significant, then PH assumption is violated. It should be noted that different functions of time can be used for different covariates (Kleinbaum and Klein, 2005). The use of this method is also suggested by other researchers (*e.g.,* Lee and Wang, 2003; and Kleinbaum and Klein, 2005).

Therneau and Grambsch (2000) point out that even though the plots and tests of PH based on the coefficients of the interaction of time-invariant covariates and some function of time are powerful, they may fail to detect some forms of non-PH.

Thus, a Cox model with all the covariates in study and the interactions of these covariates with the functions of time t and $\ln(t)$ was estimated. The results of these models are shown on Table 10. Only the estimates of the interactions are presented.

Table 10 – Estimates of the model with interaction of time-invariant covariates and the functions of time t and $\ln(t)$ of fixed-telephone contracts

	Function $\_t$		Function $\ln(\_t)$	
	β	***p*-value**	β	***p*-value**
N_total_dunning	0.0019	0.000**	0.440	0.000**
Mean_overall_revenues	0.0000	0.000**	0.001	0.061
Current_debts	-0.0001	0.000**	-0.016	0.000**
Mean_int_out_value	0.0000	0.198	-0.005	0.015*
Mean_int_in_duration	0.0000	0.924	0.000	0.346
Mean_loc_out_value	0.0003	0.000**	0.198	0.000**
Mean_loc_in_duration	0.0000	0.193	0.000	0.676
Mean_nat_out_value	0.0001	0.213	0.023	0.593
Mean_nat_in_duration	0.0000	0.602	0.000	0.527
Mean_mobile_value	0.0000	0.234	-0.007	0.021*
Mean_other_value	0.0000	0.045*	0.002	0.729
Mean_other_duration	0.0000	0.003**	-0.001	0.000**
Mean_quantity_calls_out	0.0000	0.008**	0.004	0.001**
Portability	0.0036	0.000**	1.614	0.000**
Payment_method	-0.0020	0.000**	-0.660	0.000**
Flat_plan_teleph_1	0.0017	0.000**	1.145	0.000**
Flat_plan_teleph_2	0.0008	0.003**	0.494	0.005**
Flat_plan_teleph_3	0.0018	0.000**	1.220	0.000**
Equipment_renting	-0.0002	0.450	0.147	0.143
Beja	0.0010	0.260	0.506	0.280
Braga	0.0005	0.110	-0.093	0.379
Castelo Branco	0.0004	0.628	0.269	0.521
Coimbra	0.0006	0.074	0.149	0.257
Évora	0.0010	0.351	0.194	0.622
Faro	0.0009	0.014*	0.284	0.025*
Guarda	0.0009	0.371	0.155	0.714
Leiria	0.0007	0.092	0.024	0.874
Lisboa	0.0006	0.016*	0.188	0.017*
Portalegre	0.0004	0.617	0.170	0.515
Porto	0.0000	0.899	-0.195	0.014*
Santarém	0.0010	0.009**	0.275	0.069
Setúbal	0.0006	0.040*	0.028	0.770
Viana do Castelo	0.0002	0.639	0.064	0.733
Vila Real	-0.0009	0.160	-0.509	0.006**
Viseu	-0.0004	0.612	-0.127	0.620
Gender	0.0003	0.006**	0.110	0.004**

** significant at the 1% level; * significant at the 5% level

In both models, almost 50 percent of the interactions are statistically significant, which allows concluding that the PH assumption is violated for these covariates, and, thus, the model is not PH on the whole.

4.3.3.5. Comparing the fitting of PH and non-PH models

Lee and Wang (2003) propose a comparison of the goodness-of-fit of PH and non-PH models, and they argue that if non-PH parametric models provide a better fit than PH models, the data do not satisfy the PH assumption. Bradburn *et al.* (2003) also point out that the model that best fits the data is the most appropriate model to be used (between PH and non-PH).

In order to compare the fitting of the PH and non-PH models, the AIC and the BIC were computed for the exponential, Weibull (PH models), lognormal, and log-logistic (AFT models) models. The results are presented in Table 11.

Table 11 – AIC and BIC of the PH and AFT models of fixed-telephone contracts

	Exponential	**Weibull**	**Lognormal**	**Log-logistic**
AIC	39 472.30	35 161.09	37 039.02	33 822.59
BIC	39 629.96	35 346.57	37 326.52	34 045.17
df	17	20	31	24

As can be seen in the above table, the log-logistic model produces the lowest value of the AIC and BIC, which means that the model that best fits the data is the log-logistic. As such, this AFT model outperforms the PH models, which indicates that the data is not PH.

4.3.3.6. Conclusion about the PH assumption

Table 12 presents a summary of the PH assumption tests presented above. From its analysis, it can be concluded that the PH assumption fails for many variables, at most in one type of test. Moreover, not all tests generate the same conclusion for the covariates.

Table 12 – Summary of the PH assumption tests of fixed-telephone contracts

	Piecewise regressions	**Statistical tests**	**Graphical (Schoenfeld residuals)**	**Graphical approach (other)**[6]	**Interaction of TIC and *_t***	**Interaction of TIC and *ln(_t)***
N_total_dunning	x	x			x	x
Mean_overall_revenues	x	x			x	
Current_debts	x	x	x		x	x
Mean_int_out_value	x	x				x
Mean_int_in_duration	x					
Mean_loc_out_value		x			x	x
Mean_loc_in_duration						
Mean_nat_out_value	x					
Mean_nat_in_duration	x					
Mean_mobile_value						x
Mean_other_value		x			x	
Mean_other_duration	x				x	x
Mean_quantity_calls_out	x				x	x
Portability	x	x	x	x	x	x
Payment_method	x	x	x	x	x	x
Flat_plan_teleph_1	x	x	x	x	x	x
Flat_plan_teleph_2	x	x		x	x	x
Flat_plan_teleph_3	x	x	x	x	x	x
Equipment_renting						
Beja		x		*		
Braga	x	x		*		
Castelo Branco				*		
Coimbra		x		*		
Évora				*		
Faro		x		*	x	x
Guarda				*		
Leiria	x	x		*		
Lisboa		x		*	x	x
Portalegre				*		
Porto	x			*		x
Santarém	x	x		*	x	
Setúbal		x		*	x	
Viana do Castelo	x			*		
Vila Real				*		x
Viseu				*		
Gender		x			x	x

x – PH assumption fails or seems to fail
* – the graphic analysis does not allow any conclusion

[6] $-\ln\left\{-\ln\left[\hat{S}(t)\right]\right\}$ against $\ln(t)$

4.3.4. Model estimation

As mentioned above, it seems that the model that best fits the data is the log-logistic, because it produces the lowest AIC.

In order to test for the presence of unobserved individual heterogeneity, a log-logistic model with gamma-distributed frailty (unshared) was estimated. There is statistical evidence of unobserved individual heterogeneity ($H_0 : \theta = 0;\ p < 0.000$), and thus, this effect has to be included in the model because it improves the results. The final model is presented in Table 13.

Table 13 - Estimates of the log-logistic model with gamma-distributed unshared frailty of fixed-telephone contracts

	Mean/ proportion	Log-logistic (gamma frailty)		
		β	Std. error	*p*-value
N_total_dunning	0.08	-1.462	0.025	0.000**
Mean_overall_revenues	42.93	-0.026	0.001	0.000**
Current_debts	35.55	0.022	0.000	0.000**
Mean_int_out_value	0.99	0.021	0.002	0.000**
Mean_int_in_duration	13.02	0.001	0.000	0.000**
Mean_loc_out_value	0.97	0.039	0.004	0.000**
Mean_loc_in_duration	96.36	0.000	0.000	0.000**
Mean_nat_out_value	0.21	0.039	0.011	0.001**
Mean_nat_in_duration	35.07	0.000	0.000	0.004**
Mean_mobile_value	2.61	0.020	0.002	0.000**
Mean_other_value	0.94	-0.040	0.004	0.000**
Mean_other_duration	15.34	-0.001	0.000	0.000**
Portability	0.14	0.174	0.017	0.000**
Payment_method	0.77	-0.080	0.017	0.000**
Flat_plan_teleph_1	0.02	0.377	0.037	0.000**
Flat_plan_teleph_2	0.03	0.336	0.037	0.000**
Flat_plan_teleph_3	0.06	0.112	0.023	0.000**
Equipment_renting	0.02	-0.143	0.044	0.001**
Braga	0.05	0.178	0.028	0.000**
Lisboa	0.41	0.166	0.015	0.000**
Porto	0.27	0.080	0.016	0.000**
Constant		8.154	0.029	0.000**
ln gamma		-0.9996	0.012	0.000**
ln theta		-0.716	0.094	0.000**
gamma		0.368	0.004	
theta		0.489	0.046	

** significant at the 1% level; * significant at the 5% level

Our results show that customers that spend more on international, national, local, and mobile calls outside the pack, have larger survival times. Moreover, customers that spend more time on international, national, and local calls inside the pack also have a larger contract lifetime. Both indicators are related to the fixed-telephone usage; but whereas the first group of indicators has no additional cost for the customer, the second group has. As such, it seems that customers with harder usage of the fixed-telephone service have a longer relationship with the service provider. Among these variables, those that have a greater influence on contract lifetime are the mean value spent on local and national calls outside the pack. On the other hand, there is empirical evidence that the time and money spent on other type of calls (*e.g.*, value-added calls, special numbers, etc.) negatively affect the survival time of the fixed-telephone contracts.

The results of the present study also indicate that customers with greater average monthly spending with the service provider have shorter contract lifetimes. This result is consistent with the results of Zhang *et al.* (2006), who found that the overall revenues from the last 6 months negatively affects the survival time of contracts in the fixed-telephone industry.

Moreover, it seems that the total number of overdue bills (since ever) negatively affect the survival time. Thus, for each additional invoice in debt, the contract lifetime reduces about 78%. In fact, this is the variable with greater impact on fixed-telephone contract lifetime.

Contrary to expectations, it seems that the value of current debts of the customer has a positive effect on survival time. This can be due to the fact that, until recently, the firm's policy was not stopping the service to customers with debts.

The results of the present study also indicate that the survival time for customers that required portability is larger than for those that did not require portability. Contracts paid by direct debit also last longer than contracts paid by other methods. Zhang *et al.* (2006) also found that the probability of churn increases for more difficult payment methods. There is also empirical evidence that the fixed-telephone contracts with one of the available flat plans have larger survival times than those without one of the flat plans. Furthermore, the contracts of those customers who buy the equipment for the installation of the service have longer survival times than those of customers who rent the equipment.

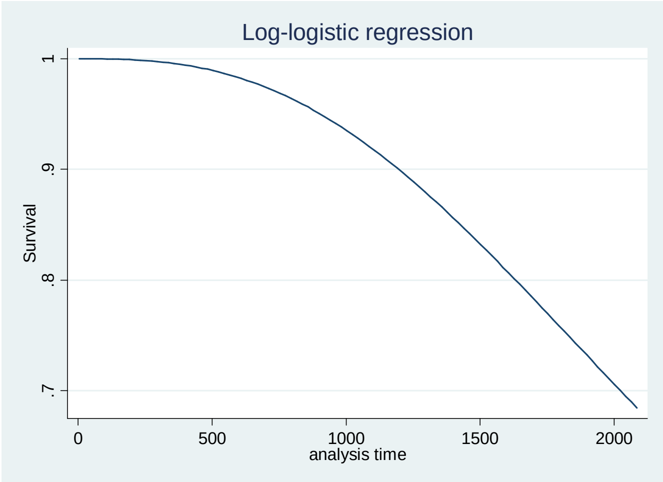
The results of the model appear to indicate that contracts of customers from the provinces of Braga, Lisboa, and Porto last longer than those of customers from Aveiro.

Lastly, the results suggest that the customer retention rate (as regards to the fixed-telephone contracts) is neither constant over time (because the exponential model is the only one which hazard function is constant and this model does not definitely adequately fits the data) nor across customers (because the PH assumption is not satisfied), which contradicts a common assumption made by several researchers on the CLV computation, in section 2.5.

The hazard and survival curves are presented in Figures 15 and 16, respectively. As can be seen from the analysis of the population hazard curve, there is duration dependence. In fact, the probability that a customer cancels a fixed-telephone contract with the service provider increases as the customer lifetime increases. Zhang *et al.* (2006) also found the existence of duration dependence on fixed-telephone contracts.

Figure 15 – Hazard function of fixed-telephone contracts

Figure 16 – Survival function of fixed-telephone contracts

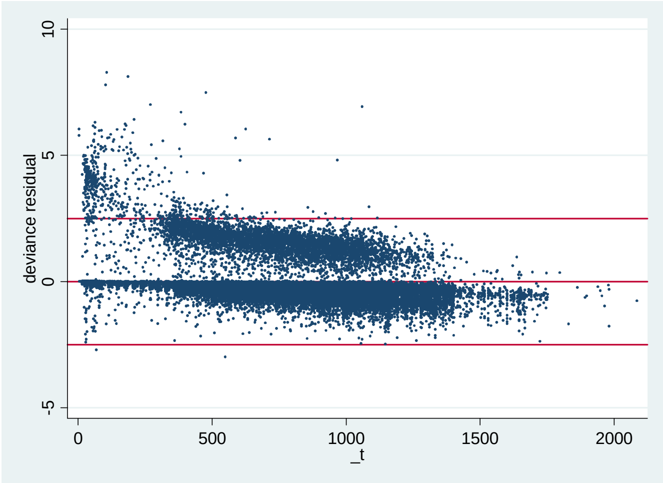


4.3.5. Identification of outliers

Box-Steffensmeier and Jones (2004) propose to identify the outlier observations by plotting the deviance residuals against the observation number or against survival time (Box-Steffensmeier and Jones, 2004). They also suggest plotting the smoothed residuals to help the visualization. Whereas the plot of the deviance residuals against the observation number allows to easily identify the observations with large residuals (which are potential outliers), the plot of the deviance residuals against survival time also allows to conclude if there are an apparent relationship between the positive/negative deviance residuals and time (*i.e.*, for instance, if large negative deviance residuals are concentrated in longer survival times) (Box-Steffensmeier and Jones, 2004).

As suggested by Box-Steffensmeier and Jones (2004), in order to identify the outliers, the deviance residuals were plotted against the survival time (Figure 17). The graph shows that the majority of outliers are concentrated on the lowest survival times. Nevertheless, only about 0.73% of the observations are outliers.

Figure 17 – Deviance residuals of the model of fixed-telephone contracts



4.3.6. Analysis of the goodness-of-fit of the model

As mentioned in section 3.8, the Cox-Snell residuals are useful to examine the goodness-of-fit of the model. An important property of these residuals is that if the selected model adequately fits the data, the Cox-Snell residuals have an unit exponential distribution[7] (Box-Steffensmeier and Jones, 2004; Cleves *et al.*, 2004; Klein and Moeschberger, 1997; Lee and Wang, 2003) with density function defined as (Lee and Wang, 2003)

$$f\left(r_{cs}\right) = e^{-r_{cs}} \qquad \textbf{(118),}$$

and survival function defined as (Lee and Wang, 2003)

$$S\left(r_{cs}\right) = \int_{r_{cs}}^{\infty} f\left(x\right)dx = \int_{r_{cs}}^{\infty} e^{-x}dx = e^{-r_{cs}} \qquad \textbf{(119).}$$

Let $\hat{S}\left(r_{cs_i}\right)$ be a Kaplan-Meier estimate and $\hat{H}\left(r_{cs_i}\right)$ a Nelson-Aalen estimate. If the fitted model is appropriate, the plot of r_{cs_i} against $\hat{H}\left(r_{cs_i}\right)$ is a straight line with slope 1 and

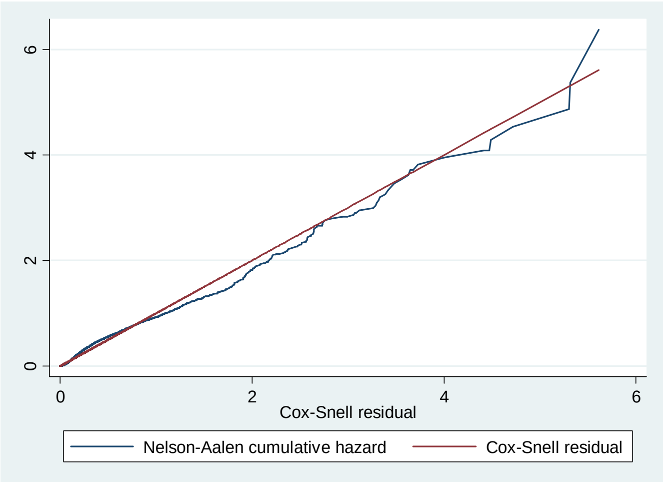
[7] Collet (1994) points out that this property may not be satisfied in small samples.

zero intercept (Blossfeld *et al.*, 2007; Box-Steffensmeier and Jones, 2004; Cleves *et al.*, 2004; Collet, 1994; Lee and Wang, 2003).

Even though some deviations from the reference line of 45^0 may be expected, mainly at the end of the integrated hazard function (due to the reduced number of individuals in the sample), systematic deviations from the reference line of 45^0 may indicate lack of fit of the model (Box-Steffensmeier and Jones, 2004; Cleves *et al.*, 2004; Collet, 1994).

The goodness of fit of the model is tested by plotting the Nelson-Aalen cumulative hazard estimator for Cox-Snell residuals, which is presented in Figure 18. From the analysis of this graph, it can be concluded that the model adequately fits the data, because the plot shows a line with slope approximately equal to one.

Figure 18 - Cumulative hazard of Cox-Snell residuals of the model of fixed-telephone contracts



The goodness-of-fit of the model can be also evaluated using the AIC and BIC, as explained in section 3.5.4.8. Table 14 presents a summary of some measures of goodness-of-fit, like the log-likelihood of the null model and the final model, the AIC, and the BIC. As mentioned above, this model produces the lowest AIC among all tested distributions.

Table 14 – Some statistics to measure the goodness-of-fit of the model of fixed-telephone contracts

	Log-logistic (gamma frailty)
Log-likelihood (null)	-22 082.16
Log-likelihood (model)	-16 819.16
df	27
AIC	33 692.33
BIC	33 942.73

4.4. Partial customer churn: The ADSL contracts

4.4.1. Selection of covariates

The selection of covariates to be included in the model took into consideration the correlation of the covariates in order to avoid the inclusion of correlated covariates. Table 15 presents the selected covariates to be used in the hazard model of the ADSL service.

Table 15 - Selected covariates to the hazard model of ADSL contracts

		Covariates
Contract	1	Portability
	2	Payment_method
	3	Flat_plan_ADSL_1
	4	Equipment_renting
Customer demographics	5	Gender
	6	Beja
	7	Braga
	8	Castelo Branco
	9	Coimbra
	10	Évora
	11	Faro
	12	Guarda
	13	Leiria
	14	Lisboa
	15	Portalegre
	16	Porto
	17	Santarém
	18	Setúbal
	19	Viana do Castelo
	20	Vila Real
	21	Viseu

		Covariates
Payment history	22	N_total_dunning
	23	Current_debts
Global revenues	24	Mean_overall_revenues
ADSL revenue	25	Mean_value_additional_traffic
ADSL usage	26	Mean_internet_traffic

4.4.2. Analysis of the functional form of covariates

As presented for the fixed-telephone model, the functional form of the covariates is examined by the analysis of the plots of the smoothed martingale residuals against each continuous covariate. The graphs are shown in Appendix E. From the analysis of these graphs, it can be concluded that the plots are approximately linear, and, thus, no known transformation id required.

4.4.3. Testing the PH assumption

As presented for the fixed-telephone contracts, the PH assumption is also analysed by using piecewise regressions, statistical tests and graphical approaches, tests for the coefficients for the interaction of time-invariant covariates and a function of time, and lastly, comparing the fitting of PH and non-PH models.

4.4.3.1. Piecewise regression

As mentioned for the fixed-telephone contracts, the database was divided into two groups. The first group includes the contracts whose lifetime is less than or equal to the median lifetime and the second group includes the remaining contracts. The median lifetime of ADSL contracts is 783 days. The models are presented in Table 16.

As can be observed in Table 16, the coefficients of some covariates are not consistent across the two groups. Moreover, the estimated coefficients of some covariates present an opposite sign between the two groups in study. This indicates that the PH assumption is not satisfied, as mentioned by Box-Steffensmeier and Zorn (2001) and Box-Steffensmeier and Jones (2004).

Table 16 – Estimates of the piecewise Cox models of ADSL contracts

	Group 1		Group 2	
	β	***p*-value**	β	***p*-value**
N_total_dunning	1.823	0.000**	3.019	0.000**
Mean_overall_revenues	0.018	0.000**	0.011	0.000**
Current_debts	-0.028	0.000**	-0.083	0.000**
Mean_internet_traffic	0.000	0.041*	0.000	0.014*
Mean_value_additional_traffic	-0.091	0.000**	0.026	0.106
Payment_method	0.382	0.000**	-0.503	0.000**
Equipment_renting	0.263	0.012*	0.275	0.247
Flat_plan_ADSL_1	0.162	0.003**	-0.324	0.000**
Beja	-0.406	0.179	0.344	0.424
Braga	-0.098	0.418	-0.411	0.009**
Castelo Branco	-0.368	0.192	-0.610	0.157
Coimbra	-0.225	0.066	0.014	0.939
Évora	-0.492	0.134	-0.359	0.618
Faro	-0.055	0.658	0.018	0.930
Guarda	-0.336	0.266	0.063	0.904
Leiria	-0.375	0.008**	-0.069	0.737
Lisboa	-0.307	0.001**	-0.316	0.025*
Portalegre	-0.258	0.198	-0.697	0.333
Porto	-0.063	0.503	-0.310	0.029*
Santarém	-0.434	0.000**	0.166	0.404
Setúbal	-0.090	0.403	0.109	0.492
Viana do Castelo	-0.507	0.002**	-0.367	0.142
Vila Real	0.041	0.847	0.135	0.709
Viseu	-0.207	0.361	-0.868	0.228
Gender	0.002	0.953	-0.014	0.750

** significant at the 1% level; * significant at the 5% level

4.4.3.2. Statistical tests based on residuals

The PH assumption was tested for each covariate with the Rao efficient score test of Therneau and Grambsch and (1994) and the PH assumption of the global model was tested using the Grambsch and Therneau (1994) test. Table 17 shows the results of both tests.

Table 17 – Statistical tests of the PH assumption of ADSL contracts

	Rho	**Chi2**	**Df**	***p*-value**
N_total_dunning	0.271	1 225.46	1	0.000**
Mean_overall_revenues	-0.033	11.38	1	0.001**
Current_debts	-0.263	3 775.85	1	0.000**
Mean_internet_traffic	0.061	21.84	1	0.000**
Mean_value_additional_traffic	-0.030	5.25	1	0.022*
Payment_method	-0.195	372.66	1	0.000**
Equipment_renting	-0.022	2.94	1	0.086
Flat_plan_ADSL_1	-0.050	15.87	1	0.000**
Beja	0.027	4.47	1	0.035*
Braga	0.026	4.61	1	0.032*
Castelo Branco	0.011	0.81	1	0.369
Coimbra	0.035	7.98	1	0.005**
Évora	0.022	2.92	1	0.087
Faro	0.038	9.78	1	0.002**
Guarda	0.014	1.16	1	0.282
Leiria	0.052	17.70	1	0.000**
Lisboa	0.048	16.06	1	0.000**
Portalegre	0.017	1.93	1	0.165
Porto	0.015	1.62	1	0.203
Santarém	0.051	17.52	1	0.000**
Setúbal	0.044	13.20	1	0.000**
Viana do Castelo	0.009	0.58	1	0.448
Vila Real	0.007	0.29	1	0.590
Viseu	0.006	0.22	1	0.636
Gender	0.024	3.59	1	0.058
Global test		4 130.61	26	0.000**

** significant at the 1% level; * significant at the 5% level

The analysis of Table 17 allows us to conclude that PH assumption is not satisfied for about 60 percent of the covariates, and, consequently, we can assume that the whole model is not PH.

4.4.3.3. Graphical approaches based on residuals

The PH assumption was also examined by the analysis of the two graphical approaches used for the fixed-telephone model. The analysis of the graphs of the Schoenfeld residuals (see Appendix F) seems to indicate that the covariates *Current_debts* and *Payment_method* may not satisfy the PH assumption.

On the other hand, according to the analysis of the graphs of the $-\ln\left\{-\ln\left[\hat{S}(t)\right]\right\}$ against $\ln(t)$ (Appendix G), it seems that the PH assumption fails in the discrete covariates *Payment_method* and *Flat_plan_ADSL_1*. Note that the graph of the covariate Province does not allow any conclusion, due to the multiplicity of categories of this covariate.

4.4.3.4. Testing the coefficients of the interaction of time-invariant covariates and the functions of time t and $\ln(t)$

The results of a Cox model with all the covariates in study and their interactions with the functions of time t and $\ln(t)$ are presented in Table 18 (note that only the interactions are shown in this table). About 40 and 50 percent of the interactions are statistically significant (in the first and second situations, respectively), which indicates that the PH assumption is not verified for these covariates. As such, the model, as a whole, is not PH.

Table 18 - Estimates of the model with interaction of time-invariant covariates and the functions of time t and $\ln(t)$ of ADSL contracts

	Function $\_t$		Function $\ln(\_t)$	
	β	***p*-value**	β	***p*-value**
N_total_dunning	0.002	0.000**	0.439	0.000**
Mean_overall_revenues	0.000	0.000**	0.000	0.717
Current_debts	0.000	0.000**	-0.016	0.000**
Mean_internet_traffic	0.000	0.000**	0.000	0.000**
Mean_value_additional_traffic	0.000	0.000**	0.323	0.000**
Payment_method	-0.002	0.000**	-0.753	0.000**
Equipment_renting	0.000	0.681	-0.002	0.982
Flat_plan_ADSL_1	0.000	0.036*	-0.002	0.973
Beja	0.001	0.194	0.526	0.292
Braga	0.000	0.241	-0.158	0.140
Castelo Branco	0.000	0.747	0.179	0.662
Coimbra	0.000	0.201	0.069	0.599
Évora	0.001	0.460	0.020	0.960
Faro	0.001	0.024*	0.287	0.023*
Guarda	0.001	0.469	0.024	0.953
Leiria	0.001	0.044*	0.050	0.752
Lisboa	0.001	0.013*	0.199	0.009**
Portalegre	0.000	0.594	0.213	0.443
Porto	0.000	0.965	-0.221	0.004**
Santarém	0.001	0.008**	0.248	0.101
Setúbal	0.000	0.118	-0.035	0.704
Viana do Castelo	0.000	0.913	-0.033	0.855
Vila Real	-0.001	0.362	-0.512	0.006**
Viseu	-0.001	0.399	-0.224	0.380
Gender	0.000	0.007**	0.121	0.001**

** significant at the 1% level; * significant at the 5% level

4.4.3.5. Comparing the fitting of PH and non-PH models

PH (exponential and Weibull) and AFT models (lognormal and log-logistic) were compared on the basis of AIC and BIC. The results are presented in Table 19. The model that best fit the data is the log-logistic, because this model produces the lowest value of AIC. Consequently, it can be concluded that the data is probably not PH as an AFT model has the best goodness of fit.

Table 19 – AIC and BIC of the PH and AFT models of ADSL contracts

	Exponential	Weibull	Lognormal	Log-logistic
AIC	39 636.74	35 407.14	37 369.84	34 262.08
BIC	39 748.03	35 536.98	37 564.6	34 382.65
df	12	14	21	13

4.4.3.6. Conclusion about the PH assumption

A summary of the PH assumption tests presented above is shown on Table 20. As shown, it can be concluded that the PH assumption fails in almost all variables, at most in one type of test. It should be emphasised that the conclusion about the PH for the covariates differs across tests.

Table 20 – Summary of the PH assumption tests of ADSL contracts

	Piecewise regressions	Statistical tests	Graphical (Schoenfeld residuals)	Graphical approach (other)[8]	Interaction of TIC and _*t*	Interaction of TIC and *ln(_t)*
N_total_dunning	x	x			x	x
Mean_overall_revenues		x			x	
Current_debts	x	x	x		x	x
Mean_internet_traffic		x			x	x
Mean_value_additional_traffic	x	x			x	x
Payment_method	x	x	x	x	x	x
Equipment_renting	x					
Flat_plan_ADSL_1	x	x		x	x	
Beja		x		*		
Braga	x	x		*		
Castelo Branco				*		
Coimbra		x		*		
Évora				*		
Faro		x		*	x	x
Guarda				*		
Leiria	x	x		*	x	
Lisboa		x		*	x	x
Portalegre				*		
Porto	x			*		x
Santarém	x	x		*	x	
Setúbal		x		*		
Viana do Castelo	x			*		
Vila Real				*		x
Viseu				*		
Gender					x	x

x – PH assumption fails or seems to fail
* – the graphic analysis does not allow any conclusion

4.4.4. Model estimation

The hazard function of the ADSL contracts is estimated using a log-logistic model, because, as mentioned above, it seems that this is the model that best fits the data, as it produces the lowest AIC.

The presence of unobserved individual heterogeneity is tested by estimating a log-logistic model with gamma-distributed frailty (unshared). There is statistical evidence of

[8] $-\ln\left\{-\ln\left[\hat{S}(t)\right]\right\}$ against $\ln(t)$

unobserved individual heterogeneity ($H_0: \theta = 0; p < 0.000$), and thus, this effect was included in the model. The final model is presented in Table 21.

Table 21 - Estimates of the log-logistic model with gamma-distributed unshared frailty of ADSL contracts

	Mean/ proportion	**Log-logistic (gamma frailty)**		
		β	**Std. error**	***p*-value**
N_total_dunning	0.08	-1.511	0.025	0.000**
Mean_overall_revenues	42.93	-0.020	0.001	0.000**
Current_debts	35.55	0.022	0.000	0.000**
Mean_value_additional_traffic	0.26	0.031	0.006	0.000**
Payment_method	0.77	-0.077	0.017	0.000**
Equipment_renting	0.02	-0.178	0.045	0.000**
Flat_plan_ADSL_1	0.13	0.114	0.017	0.000**
Braga	0.05	0.188	0.029	0.000**
Lisboa	0.41	0.176	0.015	0.000**
Porto	0.27	0.096	0.016	0.000**
Constant		8.054	0.029	0.000**
ln gamma		-0.957	0.012	0.000**
ln theta		-0.971	0.120	0.000**
gamma		0.384	0.004	
theta		0.379	0.045	

** significant at the 1% level; * significant at the 5% level

The results of the present study indicate that despite the traffic on the internet (that measures the internet usage) is not a significant covariate, the value that customers spend on additional traffic is. This suggests that customers with different levels of internet usage do not have different probabilities of churn, but customers with more additional usage have longer relationships with the service provider. Moreover, as seems to happen to the fixed-telephone contracts, ADSL contracts with flat plans have a lower risk of churn than those without flat plans.

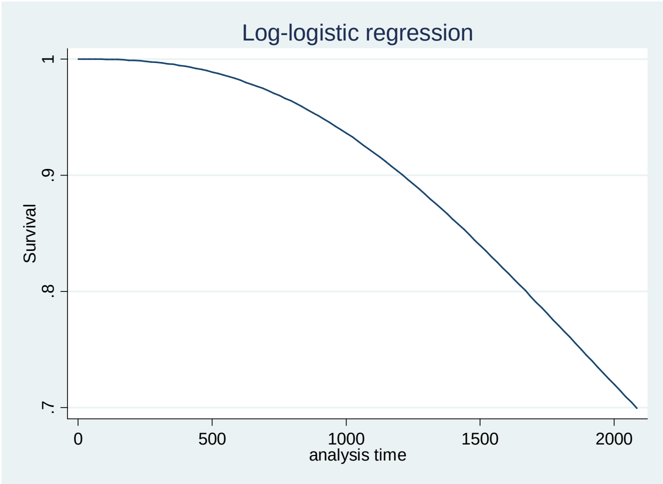
As regards to the remaining significant covariates of the model, they have a similar influence on the cancellation of ADSL contracts as verified for the fixed-telephone contracts. To the best of our knowledge, Madden *et al.* (1999) is the unique published study about customer churn on the ISP industry and they also found that the monthly spending of customers with the service provider has a positive effect on the duration of the relationship.

Lastly, the results seem to contradict a common assumption made by several researchers on the CLV computation that the customer retention rate is constant over time and across customers, as the hazard function of ADSL contracts is neither constant over time (because the exponential model is the only one for which the hazard function is constant and this model does not definitely adequately fits the data) nor across customers (because the PH assumption is not satisfied).

Figures 19 and 20 show the hazard and survival curves, respectively. The hazard curve indicates the existence of positive duration dependence because the probability that a customer cancels an ADSL contract with the service provider increases as the contract lifetime increases.

Figure 19 – Hazard curve of ADSL contracts

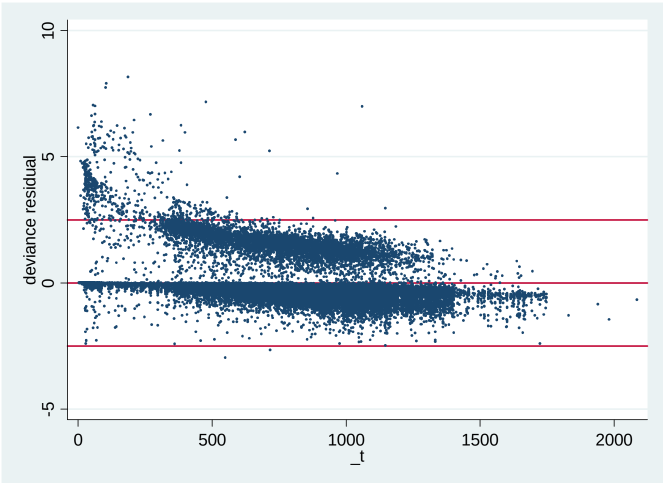
Figure 20 – Survival curve of ADSL contracts



4.4.5. Identification of outliers

Figure 21 shows the deviance residuals against the survival time, which allows to identify the outlier observations. The graph shows that the proportion of outlier observations is low (0.71%) and the majority of them are concentrated on the lowest survival times.

Figure 21 – Deviance residuals of the model of ADSL contracts



4.4.6. Analysis of the goodness-of-fit of the model

The Nelson-Aalen cumulative hazard estimator for Cox-Snell residuals indicates that the model adequately fits the data, because the line of the Nelson-Aalen cumulative hazard estimator for Cox-Snell residuals shows a slope of approximately 1 (see Figure 22).

Table 22 presents a summary of some measures of goodness of fit, like the log-likelihood of the null model and the final model, the AIC, and the BIC.

Figure 22 - Cumulative hazard of Cox-Snell residuals of the model of ADSL contracts

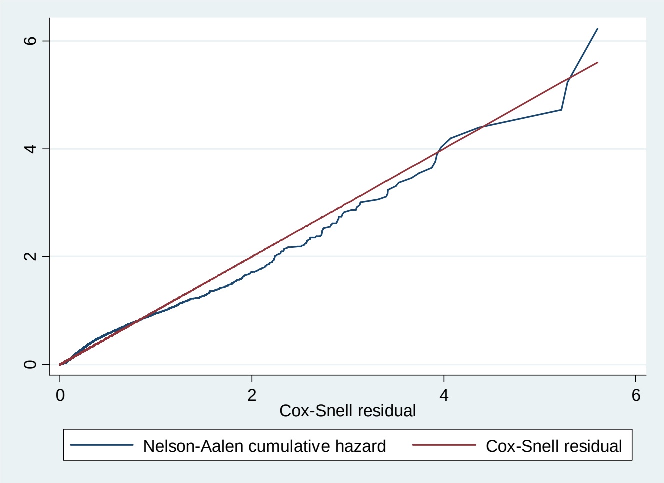


Table 22 - Some statistics to measure the goodness-of-fit of the model of ADSL contracts

	Log-logistic (gamma frailty)
Log-likelihood (null)	-21 972.42
Log-likelihood (model)	-17 118.04
df	13
AIC	34 262.08
BIC	34 382.65

4.5. Analysis of the impact of customer satisfaction on partial customer churn (fixed-telephone contracts)

Customer satisfaction has been declared as a determinant of customer retention. In fact, many researchers have argued that satisfied customers stay loyal to the product/service provider. As such, in this part of the study, our main objective is to examine the effect of the overall customer satisfaction on the cancellation of the contracts with the service provider. This empirical analysis is based on a random sample of about 700 residential customers who completed a questionnaire about customer satisfaction. The hazard model was estimated with the variable customer satisfaction and all the variables mentioned in section 4.3.1, except the (i) *Flat_plan_teleph_1*, (ii) *Flat_plan_teleph_2*, (iii) *Flat_plan_teleph_3*, (iv) *Equipment_renting*, (v) *Flat_plan_ADSL_1*, and (vi) *Province*. These discrete covariates were not included because they have categories with only few observations. Customer satisfaction is measured in a Likert scale (1 – very dissatisfied; 10 – very satisfied).

4.5.1. Analysis of the functional form of covariates

From the analysis of the plots of the smoothed martingale residuals against each continuous covariate, it can be concluded that the continuous covariates have an approximately linear behaviour, as can be seen in Appendix H.

4.5.2. Testing the PH assumption

As shown for the models presented above, the PH assumption is analysed by using piecewise regression, statistical tests and graphical approaches, tests for the coefficients for the interaction of time-invariant covariates and a function of time, and lastly, comparing the fitting of PH and non-PH models.

4.5.2.1. Piecewise regression

The database was divided into two groups. Once more, the first group includes the contracts whose lifetime is up to the median lifetime and the second group includes the remaining contracts. The median lifetime of the fixed-telephone contracts included in this sample is 843 days. The models are presented in Table 23.

Table 23 – Estimates of the piecewise models of fixed-telephone contracts (with satisfaction)

	Group 1		**Group 2**	
	β	***p*-value**	β	***p*-value**
N_total_dunning	8.165	0.000**	6.079	0.024*
Mean_overall_revenues	-0.022	0.735	0.039	0.437
Current_debts	-0.142	0.000**	-0.118	0.000**
Mean_int_out_value	-0.230	0.394	0.020	0.909
Mean_int_in_duration	0.003	0.523	-0.005	0.810
Mean_loc_out_value	-0.256	0.695	0.000	0.999
Mean_loc_in_duration	-0.008	0.015*	-0.001	0.724
Mean_nat_out_value	-1.116	0.432	-1.431	0.346
Mean_nat_in_duration	0.004	0.139	0.010	0.034*
Mean_mobile_value	0.288	0.174	0.087	0.597
Mean_other_value	0.632	0.046*	0.421	0.047*
Mean_other_duration	0.036	0.093	-0.057	0.226
Mean_quantity_calls_out	-0.041	0.531	-0.040	0.462
Portability	-6.984	0.007**	1.606	0.062
Payment_method	-2.975	0.002**	-1.035	0.298
Gender	-2.782	0.000**	-0.212	0.759
Satisfaction	0.016	0.937	-0.105	0.462

** significant at the 1% level; * significant at the 5% level

From the analysis of Table 23, it can be concluded that the significant covariates differ across groups and only three covariates are significant in both models and two coefficients of them are not consistent across the two groups.

As such, it can be said that there is empirical evidence that the effect of some covariates on the cancellation of telephone-fixed contracts is not constant over time, which means that the PH assumption is not satisfied.

4.5.2.2. Statistical tests based on residuals

Table 24 present the results of the Rao efficient score test of Therneau and Grambsch and (1994) and the Grambsch and Therneau (1994) test for the PH assumption, for each covariate and for the global model, respectively.

Table 24 – Statistical tests of the PH assumption of fixed-telephone contracts (with satisfaction)

	Rho	Chi2	Df	*p*-value
N_total_dunning	0.131	0.39	1	0.535
Mean_overall_revenues	-0.042	0.04	1	0.846
Current_debts	-0.084	0.14	1	0.710
Mean_int_out_value	0.207	1.44	1	0.231
Mean_int_in_duration	-0.050	0.10	1	0.751
Mean_loc_out_value	0.184	1.77	1	0.183
Mean_loc_in_duration	0.286	2.01	1	0.157
Mean_nat_out_value	0.070	0.20	1	0.658
Mean_nat_in_duration	0.287	2.08	1	0.150
Mean_mobile_value	-0.013	0.00	1	0.949
Mean_other_value	0.100	0.32	1	0.571
Mean_other_duration	-0.203	1.27	1	0.261
Mean_quantity_calls_out	-0.199	1.18	1	0.278
Portability	0.354	5.52	1	0.019*
Payment_method	0.110	0.51	1	0.476
Gender	0.203	1.50	1	0.221
Satisfaction	-0.123	0.33	1	0.563
Global test		18.96	17	0.331

** significant at the 1% level; * significant at the 5% level

These statistical tests provide evidence that the PH assumption does not hold for only one covariate. Moreover, according to Grambsch and Therneau (1994) test, there is empirical evidence that the global model is PH.

4.5.2.3. Graphical approaches based on residuals

The graphs of the Schoenfeld residuals (see Appendix I) provides evidence that the following covariates may not satisfy the PH assumption: *Mean_int_out_value, Mean_loc_out_value, Mean_loc_in_duration, Mean_nat_in_duration, Mean_other_duration, Portability, Payment_method, Gender,* and *Satisfaction*.

According to the analysis of the graphs of the $-\ln\left\{-\ln\left[\hat{S}(t)\right]\right\}$ against $\ln(t)$ (Appendix J), it seems that the PH assumption fails for the covariates *Portability* and *Payment_method*, but not for *Gender*.

4.5.2.4. Testing the coefficients of the interaction of time-invariant covariates and the functions of time t and $\ln(t)$

Table 25 shows the results of a Cox model with all the covariates in study and their interactions with the functions of time t and $\ln(t)$ (only the interactions are shown in this table). Only one interaction with the function of time t, and two interactions with the function of time $\ln(t)$ are statistically significant, which indicates that the PH assumption seems to hold for all covariates except *Mean_loc_out_value* and *Portability*.

Table 25 - Estimates of the model with interaction of time-invariant covariates and the functions of time t and $\ln(t)$ of fixed-telephone contracts (with satisfaction)

	Function $\_t$		Function $\ln(\_t)$	
	β	***p*-value**	β	***p*-value**
N_total_dunning	1.003	0.498	1.619	0.512
Mean_overall_revenues	1.000	0.752	-0.056	0.578
Current_debts	1.000	0.607	-0.012	0.777
Mean_int_out_value	1.000	0.431	0.387	0.387
Mean_int_in_duration	1.000	0.651	-0.007	0.586
Mean_loc_out_value	1.002	0.138	2.327	0.050*
Mean_loc_in_duration	1.000	0.123	0.009	0.091
Mean_nat_out_value	1.003	0.539	2.993	0.431
Mean_nat_in_duration	1.000	0.263	0.008	0.238
Mean_mobile_value	1.000	0.660	0.221	0.476
Mean_other_value	1.000	0.562	0.400	0.394
Mean_other_duration	1.000	0.330	-0.026	0.474
Mean_quantity_calls_out	1.000	0.276	-0.172	0.122
Portability	1.016	0.004**	12.811	0.008**
Payment_method	1.002	0.328	1.629	0.313
Gender	1.003	0.133	1.609	0.184
Satisfaction	1.000	0.758	-0.123	0.699

** significant at the 1% level; * significant at the 5% level

4.5.2.5. Comparing the fitting of PH and non-PH models

Table 26 presents a comparison of the PH and AFT models based on the AIC and BIC. The model that best fit the data is the log-logistic, because this model has the lowest value of AIC. As such, this AFT model outperforms the PH models.

Table 26 – AIC and BIC of the PH and AFT models of fixed-telephone contracts (with satisfaction)

	Exponential	Weibull	Lognormal	Log-logistic
AIC	205.61	167.03	170.73	163.54
BIC	233.12	199.12	198.24	200.22
df	6	7	6	8

4.5.2.6. Conclusion about the PH assumption

Table 27 presents a summary of the results of the PH assumption tests computed above. From its analysis, it can be concluded that the PH assumption seems to hold for almost all variables.

Table 27 – Summary of the PH assumption tests of fixed-telephone contracts (with satisfaction)

	Piecewise regressions	**Statistical tests**	**Graphical (Schoenfeld residuals)**	**Graphical approach (other)[9]**	**Interaction of TIC and _*t***	**Interaction of TIC and *ln(_t)***
N_total_dunning	x					
Mean_overall_revenues						
Current_debts						
Mean_int_out_value			x			
Mean_int_in_duration						
Mean_loc_out_value			x			x
Mean_loc_in_duration	x		x			
Mean_nat_out_value						
Mean_nat_in_duration	x		x			
Mean_mobile_value						
Mean_other_value	x					
Mean_other_duration			x			
Mean_quantity_calls_out						
Portability	x	x	x	x	x	x
Payment_method	x		x	x		
Gender	x		x			
Satisfaction			x			

x – PH assumption fails or seems to fail

4.5.3. Model estimation

A log-logistic model with gamma-distributed frailty (unshared) was estimated. There is statistical evidence of unobserved individual heterogeneity ($H_0 : \theta = 0;\ p < 0.034$), and thus, this effect was included in the model. The final model is presented in Table 28.

[9] $-\ln\left\{-\ln\left[\hat{S}(t)\right]\right\}$ against $\ln(t)$

Table 28 - Estimates of the log-logistic model with gamma-distributed unshared frailty of fixed-telephone contracts (with satisfaction)

	Mean/ proportion	Log-logistic (gamma frailty)		
		β	Std. error	*p*-value
N_total_dunning	0.06	-1.563	0.265	0.000**
Current_debts	35.54	0.029	0.005	0.000**
Mean_other_value	0.91	-0.123	0.042	0.003**
Portability	0.15	0.419	0.166	0.011*
Gender	0.71	0.432	0.127	0.001**
Constant		6.826	0.160	0.000**
ln gamma		-1.485	0.183	0.000**
ln theta		0.680	0.681	0.318
gamma		0.226	0.042	
theta		1.974	1.345	

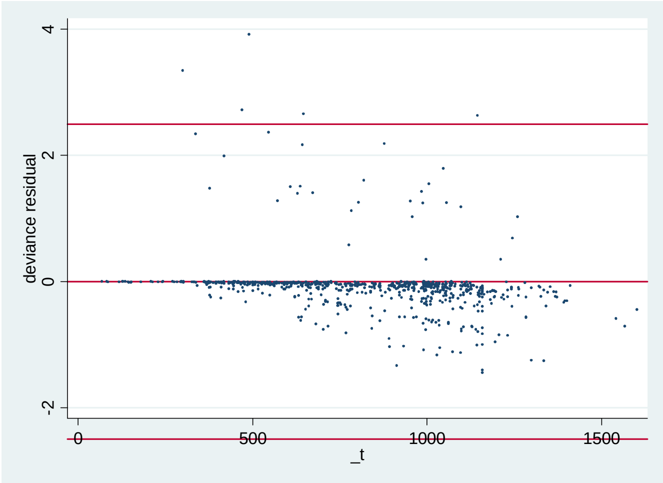
** significant at the 1% level; * significant at the 5% level

As can be seen in Table 28, customer satisfaction is not a significant covariate, which suggests that customer satisfaction in this context is not a reason for contract cancellation. A possible explanation of this finding is that even though the customer is not satisfied, he/she may do not switch to other operator due to inertia or habit. This contradicts the majority of the literature about satisfaction (*e.g.*, Bolton, 1998; Eshghi *et al.*, 2007). Kim and Yoon (2004) found that whereas some types of satisfaction positively affect the survival time, others do not have any influence, in the mobile phone industry. Van den Poel and Larivière (2004) show that some studies also did not find any influence of satisfaction on survival time.

4.5.4. Identification of outliers

The graph of the deviance residuals against survival time shows that very few observations are outliers (Figure 23).

Figure 23 – Deviance residuals of the model of fixed-telephone contracts (with satisfaction)



4.5.5. Analysis of the goodness-of-fit of the model

The plot of the Nelson-Aalen cumulative hazard estimator for Cox-Snell residuals shows only some deviations from the reference line of 45^0 and mainly at the end of the cumulative hazard function (Figure 24). As explained by Cleves *et al.* (2004), this type of deviations may be expected, and, thus, it seems that the model adequately fits the data.

A summary of some measures of goodness of fit are presented in Table 29.

Figure 24 - Cumulative hazard of Cox-Snell residuals of the model of fixed-telephone contracts (with satisfaction)

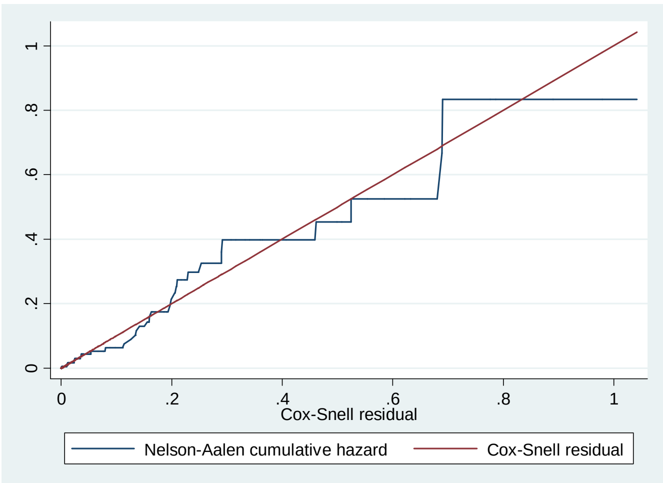


Table 29 - Some statistics to measure the goodness-of-fit of the model of fixed-telephone contracts (with satisfaction)

	Log-logistic
Log-likelihood (null)	-115.58
Log-likelihood (model)	-73.77
df	8
AIC	163.54
BIC	200.22

4.6. Analysis of the impact of customer satisfaction on partial customer churn (ADSL contracts)

4.6.1. Analysis of the functional form of covariates

The analysis of the plots of the smoothed martingale residuals against each continuous covariate allows to conclude that the continuous covariates have approximately a linear behaviour, as can be seen in Appendix K.

4.6.2. Testing the PH assumption

Once more, the PH assumption is analysed by using piecewise regressions, statistical tests and graphical approaches, tests for the coefficients for the interaction of time-invariant covariates and a function of time, and lastly, comparing the fitting of PH and non-PH models.

4.6.2.1. Piecewise regression

The database was divided into two groups. Again, the first group includes the contracts whose lifetime last up to the median lifetime (inclusive) and the second group includes the remaining contracts. The median lifetime of the ADSL contracts included in this sample is 843 days. Table 30 shows the results of the models. From its analysis, it can be concluded that the significant covariates differ across groups. Furthermore, only two covariates are significant in both models and one of them presents coefficients that are not consistent across the two groups. Thus, it seems that the effect of some covariates on the cancellation of ADSL contracts is not constant over time.

Table 30 – Estimates of the piecewise models of ADSL contracts (with satisfaction)

	Group 1		Group 2	
	β	***p*-value**	β	***p*-value**
N_total_dunning	6.212	0.000**	5.498	0.002**
Mean_overall_revenues	-0.013	0.573	0.035	0.150
Current_debts	-0.121	0.000**	-0.100	0.000**
Mean_internet_traffic	0.000	0.148	0.000	0.631
Mean_value_additional_traffic	0.862	0.007**	0.125	0.647
Payment_method	-1.316	0.058	-0.723	0.443
Gender	-1.417	0.010**	-0.574	0.350
Satisfaction	-0.050	0.715	-0.092	0.477

** significant at the 1% level; * significant at the 5% level

4.6.2.2. Statistical tests based on residuals

Table 31 presents the results of the Rao efficient score test of Therneau and Grambsch and (1994) and the Grambsch and Therneau (1994) test for the PH assumption, for each covariate and for the global model, respectively. These statistical tests provide evidence that the PH assumption holds for all covariates. Moreover, according to Grambsch and Therneau (1994) test, there is empirical evidence that the global model is PH.

Table 31 – Statistical tests of the PH assumption of ADSL contracts (with satisfaction)

	Rho	Chi2	Df	*p*-value
N_total_dunning	0.061	0.09	1	0.759
Mean_overall_revenues	0.005	0.00	1	0.987
Current_debts	-0.008	0.00	1	0.968
Mean_internet_traffic	0.018	0.01	1	0.929
Mean_value_additional_traffic	-0.088	0.19	1	0.664
Payment_method	0.165	0.99	1	0.319
Gender	0.180	1.24	1	0.266
Satisfaction	-0.170	0.72	1	0.396
Global test		4.09	8	0.849

** significant at the 1% level; * significant at the 5% level

4.6.2.3. Graphical approaches based on residuals

The analysis of the graphs of the Schoenfeld residuals (see Appendix L) seems to indicate that the covariates *Payment_method*, *Gender*, and *Satisfaction* may not satisfy the PH assumption. According to the analysis of the graphs of the $-\ln\left\{-\ln\left[\hat{S}(t)\right]\right\}$ against $\ln(t)$ (Appendix M), it seems that the PH assumption only fails for the covariate *Payment_method*.

4.6.2.4. Testing the coefficients of the interaction of time-invariant covariates and the functions of time t and $\ln(t)$

The results of a Cox model including all the covariates in study and their interactions with the functions of time t and $\ln(t)$ are presented in Table 32 (only the interactions are shown in this table). None interaction is statistically significant, which indicates that the PH assumption may hold for all covariates.

Table 32 - Estimates of the model with interaction of time-invariant covariates and the functions of time t and $\ln(t)$ of ADSL contracts (with satisfaction)

	Function $\_t$		Function $\ln(\_t)$	
	β	***p*-value**	β	***p*-value**
N_total_dunning	1.002	0.546	1.311	0.560
Mean_overall_revenues	1.000	0.924	-0.009	0.770
Current_debts	1.000	0.906	-0.005	0.891
Mean_internet_traffic	1.000	0.945	0.000	0.795
Mean_value_additional_traffic	1.000	0.622	-0.243	0.596
Payment_method	1.002	0.379	0.871	0.525
Gender	1.002	0.291	0.882	0.378
Satisfaction	1.000	0.445	-0.223	0.365

** significant at the 1% level; * significant at the 5% level

4.6.2.5. Comparing the fitting of PH and non-PH models

Based on the AIC computed for PH and AFT models, it seems that the model that best fits the data is the log-logistic, because this model has the lowest value of AIC (see Table 33). Thus, there is empirical evidence that the model is not PH.

Table 33 – AIC and BIC of the PH and AFT models of ADSL contracts (with satisfaction)

	Exponential	Weibull	Lognormal	Log-logistic
AIC	205.58	164.74	168.70	161.51
BIC	233.09	192.25	200.79	202.78
df	4	6	5	6

4.6.2.6. Conclusion about the PH assumption

Table 34 presents a summary of the PH assumption tests shown above. From its analysis, it can be concluded that the PH assumption holds in almost all variables.

Table 34 – Summary of the PH assumption tests of ADSL contracts (with satisfaction)

	Piecewise regressions	**Statistical tests**	**Graphical (Schoenfeld residuals)**	**Graphical approach (other)[10]**	**Interaction of TIC and _*t***	**Interaction of TIC and *ln(_t)***
N_total_dunning						
Mean_overall_revenues						
Current_debts						
Mean_internet_traffic						
Mean_value_additional_traffic	x					
Payment_method			x	x		
Gender	x		x			
Satisfaction			x			

x – PH assumption fails or seems to fail

4.6.3. Model estimation

As mentioned above, it seems that the model that best fits the data is the log-logistic, because it has the lowest AIC.

In order to test for the presence of unobserved individual heterogeneity, a log-logistic model with gamma-distributed frailty (unshared) was estimated. There is statistical evidence that the covariates included in the model correctly explain the behaviour of the sample and the unobserved individual heterogeneity is not presented ($H_0 : \theta = 0;\ p = 0.018$). The final model is presented in the Table 35.

Customer satisfaction is not a significant covariate, which suggests that customer satisfaction in this context is not a reason for contract cancellation, as explained for the fixed-telephone contracts.

[10] $-\ln\left\{-\ln\left[\hat{S}(t)\right]\right\}$ against $\ln(t)$

Table 35 - Estimates of the log-logistic model with gamma-distributed unshared frailty of ADSL contracts (with satisfaction)

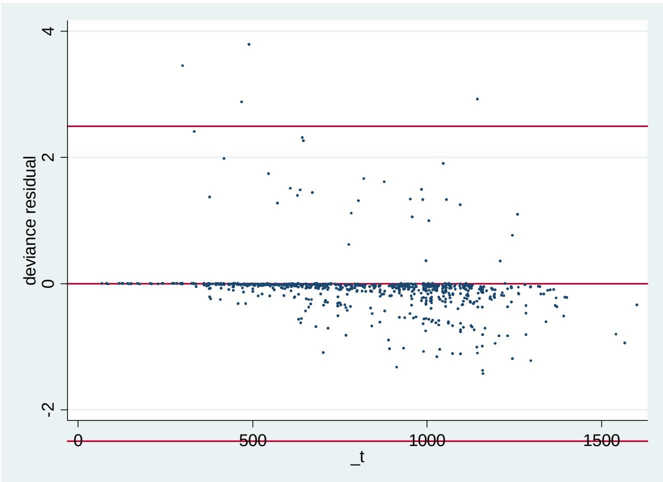
	Mean/ proportion	Log-logistic (gamma frailty)		
		β	Std. error	*p*-value
N_total_dunning	0.06	-1.431	0.288	0.000**
Current_debts	35.54	0.025	0.005	0.000**
Gender	0.71	0.282	0.123	0.022*
Constant		7.102	0.103	0.000**
ln gamma		-1.540	0.191	0.000**
ln theta		0.847	0.606	0.162
gamma		0.214	0.041	
theta		2.332	1.412	

** significant at the 1% level; * significant at the 5% level

4.6.4. Identification of outliers

Figure 25 shows the deviance residuals against survival time. The graph shows that very few observations are outliers.

Figure 25 – Deviance residuals of the model of ADSL contracts (with satisfaction)



4.6.5. Analysis of the goodness-of-fit of the model

The plot of the Nelson-Aalen cumulative hazard estimator for Cox-Snell residuals shows only some deviations from the reference line of 45^0 and mainly at the end of the cumulative hazard function (Figure 26). As explained by Cleves *et al.* (2004), this type of deviations may be expected, and, thus, it seems that the model adequately fits the data.

Table 36 presents a summary of some measures of goodness of fit, like the log-likelihood of the null model and the final model, the AIC, and the BIC. As mentioned above, this model produces the lowest AIC among all tested distributions.

Figure 26 - Cumulative hazard of Cox-Snell residuals of the model of ADSL contracts (with satisfaction)

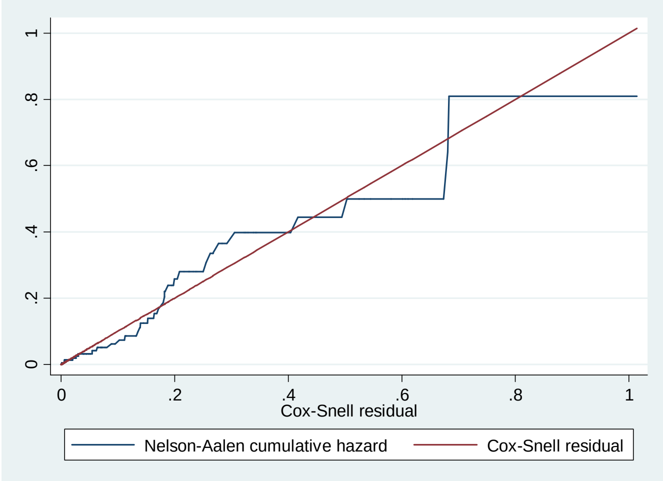


Table 36 - Some statistics to measure the goodness-of-fit of the model of ADSL contracts (with satisfaction)

	Log-logistic
Log-likelihood (null)	-115.47
Log-likelihood (model)	-71.76
df	6
AIC	161.51
BIC	202.78

4.7. Summary of the chapter

In this chapter, the hazard function of the fixed-telephone and ADSL contracts of customers from a Portuguese fixed telecommunications firm were estimated. The functional form of covariates was examined by the analysis of the plot of the smoothed curve of the martingale residuals from a null model against the values of each continuous covariate. The PH assumption is verified by using piecewise regressions, statistical tests and graphical approaches, tests for the coefficients for the interaction of time-invariant covariates and a function of time, and lastly, comparing the fitting of PH and non-PH models. Outlier observations were identified by the analysis of the plots of the deviance residuals against the survival time. Lastly, the goodness of fit of the models was examined by the analysis of the plot of the cumulative hazard of Cox-Snell residuals.

The influence of customer satisfaction on the hazard function of the fixed-telephone and ADSL contracts was also tested in this chapter.

5. CONCLUSIONS

This study sheds new light on the crucial issue of the determinants of partial customer churn in the fixed telecommunications industry in Portugal, as well as on the behaviour of the probability of partial customer churn over time and across individuals, in fixed-telephone and ADSL contracts. Considering that it is crucial to prevent the churn of potentially profitable contracts of customers in order to ensure the financial performance of the firms, the results of this study may be very valuable mainly when complemented with an analysis of the CLV for each individual.

Our results demonstrate that customers with harder usage of the fixed-telephone service have a longer relationship with the service provider. As regards to the ADSL contracts, the results provide evidence that the probability of churn does not vary with the internet usage, but customers with more additional usage than those contracted have longer relationships with the service provider. Moreover, it seems that both types of contracts with flat plans have a lower risk of churn than those without flat plans. The results of this study also indicate that customers with greater average monthly spending with the service provider have shorter contract lifetimes of both types. Moreover, it seems that the total number of overdue bills (since ever) negatively affect the survival time of both kind of contracts in study. It also seems that the survival time of fixed-telephone contracts of customers that required portability is larger than the one that did not require portability. Contracts paid by direct debit also last longer than contracts paid by other methods. Furthermore, the contracts of those customers who buy the necessary equipment last longer than those of customers who rent the equipment. The results of the model appear to indicate that the probability of churn varies across some provinces.

The results also suggest that the customer retention rate is neither constant over time nor across customers, for fixed-telephone and ADSL contracts. As such, positive duration dependence is presented in the hazard function of both types of contracts.

Contrary to our expectations, it seems that satisfaction does not influence the cancellation of both types of contracts.

Lastly, it seems that unobserved heterogeneity has an important effect in modelling the hazard function of both types of contracts.

These results have a number of managerial implications. Firstly, firms cannot make decisions about customer management based on the average churn rates. Secondly, firms must frequently estimate the probability of customer churn because this market is very dynamic. Furthermore, it appears that firms should concentrate less on customer satisfaction because it does not seem to be an important reason of customer churn, and instead focus on pricing strategy as customers appear to be sensitive to the monthly average of bills.

This study has two main limitations. Firstly, there are many other variables that might be important to be included in the models for estimating the hazard functions of the fixed-telephone and ADSL contracts (for instance, the subscription period of each contract, promotions, acquisition cost, contact details to and from the customer, complaints, customer satisfaction, other demographic data such as age, education, number of people in the household, etc). Secondly, data about usage and revenues are a static mean of a given period. The use of TVC about usage and revenues for the duration of the relationship with the customers might improve the results.

Further research should be done in order to improve the knowledge about customer churn. For instance, it would be very interesting that firms will store information about the reason for the cancellation of each contract, in order to implement competing risks models. Some researchers argue that different causes of failure exist, competing risks models produce more accurate results than two-state models. Moreover, a comparison of the behaviour of partial churn of residential and business customers will improve the state of the art in this issue. Lastly, the results of the customer churn could be included in an accurate CLV prediction in order that managers can make decisions based on rigorous models.

REFERENCES

Aalen, O.O. (1992), Modelling Heterogeneity in Survival Analysis by the Compound Poisson Distribution, *Annals of Applied Probability* 4 (2), 951-972.

Aalen, O.O., and H.K. Gjessing (2005), Stochastic Processes in Survival Analysis, *Unpublished manuscript*, http://www.google.com/books?hl=pt-PT&lr=&id=3XxUC14Y2_QC&oi=fnd&pg=PA23&dq=Stochastic+Processes+in+Survival+Analysis&ots=A_6R3KFnSY&sig=HE9EwwC8StnMVLJLakisQofkmQs#v=onepage&q=Stochastic%20Processes%20in%20Survival%20Analysis&f=false.

Ahn, J-H., S-P., Han, and Y-S. Lee (2006), Customer Churn Analysis: Churn Determinants and Mediation Effects of Partial Defection in the Korean Mobile Telecommunications Service Industry, *Telecommunications Policy* 30 (10-11), 552-568.

Akaike, H. (1974), A New Look At the Statistical Model Identification, *IEEE Transactions on Automatic Control* 19 (6), 716-723.

Allison, P. (2004). *Event History Analysis*, in Hardy M. and A. Bryman (Ed.), *Handbook of Data Analysis*, London: Sage Publications.

Andersen, P.K., and N. Keiding (2002), Multi-State Models for Event History Analysis, *Statistical Methods in Medical Research* 11 (2), 91-115.

Anderson, E.W., C. Fornell, and D.R. Lehmann (1994), Customer Satisfaction, Market Share, and Profitability: Findings fromF Sweden, *Journal of Marketing* 58 (3), 53-66.

Ang, L., and B. Taylor (2005), Managing Customer Profitability using Portfolio Matrices, *Database Marketing & Customer Strategy Management* 12 (4), 298-304.

Ata, N., and M.T. Sozer (2007), Cox Regression Models With Nonproportional Hazards Applied To Lung Cancer Survival Data, *Hacettepe Journal of Mathematics and Statistics* 36 (2), 157-167.

Bauer, C.L. (1988), A Direct Mail Customer Purchase Model, *Journal of Direct Marketing* 2 (3), 16-24.

Bauer, H.H., and M. Hammerschmidt (2005), Customer-Based Corporate Valuation, *Management Decision* 43 (3), 331-348.

Bauer, H.H., M. Hammerschmidt, and M. Braehler (2003), The Customer Lifetime Value Concept and Its Contributions to Corporate Valuation, *Yearbook of Marketing and Consumer Research* 1, 47-67.

Bayón, T., J. Gutsche, and H. Bauer (2002), Customer Equity Marketing: Touching the Intangible, *European Management Journal* 20 (3), 213-222.

Bell, D., J. Deighton, W.J. Reinartz, R.T. Rust, and G. Swartz (2002), Seven Barriers to Customer Equity Management, *Journal of Service Research* 5 (1), 77-85.

Berger, P.D., and N.I. Nasr (1998), Customer lifetime value: marketing models and applications, *Journal of Interactive Marketing* 12 (1), 17-30.

Berger, P.D., B. Weinberg, and R.C. Hanna (2003), Customer Lifetime Value Determination and Strategic Implications for a Cruise-ship Company, *Journal of Database Marketing & Customer Strategy Management* 11 (1), 40-52.

Berger, P.D., R.N. Bolton, D. Bowman, E. Briggs, V. Kumar, A. Parasuraman, and C. Terry (2002), Marketing Actions and the Value of Customer Assets: A Framework for Customer Asset Management, *Journal of Service Research* 5 (1), 39-54.

Berry, M.J.A., and G.S. Linoff (2004), *Data Mining Techniques* (2nd Edition). Indianapolis: Wiley.

Bin, L., S. Peiji, and L. Juan (2007), Customer Churn Prediction Based on the Decision Tree in Personal Handyphone System Service, *Proceedings of the International Conference on Service Systems and Service Management*.

Blattberg, R.C., and J. Deighton (1996), Manage Marketing by the Customer Equity Test, *Harvard Business Review* 74 (4), 136-144.

Blattberg, R.C., G. Getz, and J.S. Thomas (2001), *Customer Equity – Building and Managing Relationships as Valuable Assets.* Boston: Harvard Business School Press.

Blossfeld, H-P., A. Hamerle, and K.U. Mayer (1989), *Event History Analysis.* Hillsdale: Lawrence Erlbaum Associates.

Blossfeld, H-P., K. Golsch, and G. Rohwer (2007), *Event History Analysis with STATA.* USA: Lawrence Erlbaum Associates.

Bolton, R.N. (1998), A Dynamic Model of the Duration of the Customer's Relationship with a Continuous Service Provider: The Role of Satisfaction, *Marketing Science* 17 (1), 45-65.

Bolton, R.N., and C.O. Tarasi (2006), Managing Customer Relationships, *Review of Marketing Research* 3, 3-38.

Bonfrer, A., G. Knox, J. Eliashberg, and J. Chiang (2007), Diagnosing and predicting individual customer defection in a contractual setting, *Unpublished manuscript,* http://dlapp02.smu.edu.sg/getblob?blobid=1814216310067899.

Boulding, W., R. Staelin, M. Ehret, and W.J. Johnston (2005), A Customer Relationship Management Roadmap: What is Known, Potential Pitfalls, and Where to Go, *Journal of Marketing* 69 (4), 155-166.

Box-Steffensmeier, J.M., and B.S. Jones (2004), *Event History Modeling: A Guide for Social Scientists.* Cambridge: Cambridge University Press.

Box-Steffensmeier, J.M., and C. Zorn (1999), Modeling Heterogeneity in Duration Models, *Summer Meeting of the Political Methodology,* Texas A&M University.

Box-Steffensmeier, J.M., and C. Zorn (2002), Duration Models for Repeated Events, *The Journal of Politics* 64 (4), 1069-1094.

Box-Steffensmeier, J.M., and C.W. Zorn (2001), Duration Models and Proportional Hazards in Political Science, *American Journal of Political Science* 45 (4), 972-988.

Box-Steffensmeier, J.M., and S. De Boef (2002), A Monte Carlo Analysis for Recurrent Events Data, *Annual Political Methodology Meetings*, Seattle, WA.

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), 649-689.

Bradburn, M.J., T.G. Clark, S.B. Love, and D.G. Altman (2003), Survival Analysis Part II: Multivariate Data Analysis – An Introduction to Concepts and Methods, *British Journal of Cancer* 89 (3), 431-436.

Buckinx, W., and D. Van den Poel (2005), Customer Base Analysis: Partial Defection of Behaviourally Loyal Clients in a Non Contractual FMCG Retail Setting, *European Journal of Operational Research* 164 (1), 252-268.

Burez, J., and D. Van den Poel (2008), Separating Financial From Commercial Customer Churn: A Modeling Step Towards Resolving the Conflit Between the Sales and Credit Department, *Expert Systems with Applications* 35 (1-2),497-514.

Burez, J., D. Van den Poel (2007), CRM At a Pay-TV Company: Using Analytical Models to Reduce Customer Attrition by Targeted Marketing for Subscription Services, *Expert Systems with Applications* 32 (2), 277-288.

Calciu, M., and F.Salerno (2002), Customer Value Modelling: Synthesis and Extension Proposals, *Journal of Targeting, Measurement and Analysis for Marketing* 11 (2), 124-147.

Capraro, A.J., S. Broniarczyk, and R.K. Srivastava (2003), Factors Influencing the Likelihood of Customer Defection: The Role of Consumer Knowledge, *Journal of the Academy of Marketing Science* 31 (2), 164-175.

Carroll, P. (1991/92), The Fallacy of Customer Retention, *Journal of Retail Banking* 13 (4), 15-20.

Chandrashekaran, M, K. Rotte, S.S. Tax, and R. Grewal (2007), Satisfaction Strength and Customer Loyalty, *Journal of Marketing Research* 44 (1), 153-163.

Chen, Y., G. Zhang, D. Hu, and C. Fu (2007), Customer Segmentation Based on Survival Character, *Journal of Intelligent Manufacturing* 18 (4), 513–517.

Clayton, D. (1978), A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence, *Biometrika*, 65 (1) 141-151.

Cleves M. (1999), Analysis of Multiple Failure-Time Data with STATA, *STATA Technical Bulletin* 49, 30-39.

Cleves, M.A., W.W. Gould, and R.G. Gutierrez (2004), *An Introduction to Survival Analysis using STATA* (Revised Edition). Texas: Stata Press.

Colgate, M., K. Stewart, and R. Kinsella (1996), Customer Defection: A Study of the Student Market in Ireland, *The International Journal of Bank Marketing* 14 (3), 23-29.

Collet, D. (1994), *Modelling Survival Data in Medical Research.* London: Chapman & Hall.

Colombo, R., and W. Jiang (1999), A Stochastic RFM Model, *Journal of Interactive Marketing* 13 (3), 2-12.

Cox, D.R. (1972), Regression Models and Life-Tables, *Journal of the Royal Statistical Society. Series B (Methodological)* 34 (2), 187-220.

Cox, D.R., and D. Oakes (1984), *Analysis of Survival Data.* London: Chapman and Hall.

Day, G., and L. Fahey (1988), Valuing Market Strategies, *Journal of Marketing* 52 (3), 45-57.

Desai, M. (2006), Churn: A Reason to Churn!!, *Unpublished manuscript*, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=957208.

Dhar, R., and R. Glazer (2003), Hedging Customers, *Harvard Business Review* 81 (5), 86-92.

Donkers, B., P.C. Verhoef, and M. de Jong (2003), Predicting Customer Lifetime Value in Multi-Service Industries, *Working Paper*, Erasmus Research Institute of Management.

Dowling, G.R., and M. Uncles (1997), Do Customer Loyalty Programs Really Work?, *Sloan Management Review* 38 (4), 71-82.

Drew, J.H., D.R. Mani, A.L. Betz, and P.Datta (2001), Targeting Customers with Statistical and Data-Mining Techniques, *Journal of Service Research* 3 (3), 205-219.

Drèze, X., and A. Bonfrer (2005), Moving From Customer Lifetime Value to Customer Equity, *Unpublished manuscript*, http://www.xdreze.org/Publications/PBM.html.

Duboff, R.S. (1992), Marketing to Maximize Profitability, *Journal of Business Strategy* 13 (6), 10-13.

Duchateau, L., and P. Janssen (2007). *The Frailty Model.* Springer.

Dwyer, R. (1989), Customer Lifetime Valuation to Support Marketing Decision Making, *Journal of Direct Marketing* 3 (4), 8-15.

East, R., K. Hammond, and P. Gendall (2006), Fact and Fallacy in Retention Marketing, *Journal of Marketing Management* 22 (1),5-23.

Eshghi, A., D. Haughton, and H. Topi (2007), Determinants of Customer Loyalty in the Wireless Telecommunications Industry, *Telecommunications Policy* 31 (2), 93-106.

Etzion, O., A. Fisher, and S. Wasserkrug (2005), e-CLV: A Modeling Approach for Customer Lifetime Evaluation in E-Commerce Domains, With an Application and Case Study for Online Auctions, *Information Systems Frontier* 7 (4/5), 421-434.

Fader, P.S., and B.G.S. Hardie (2006), Customer-Base Valuation in a Contractual Setting: The Perils of Ignoring Heterogeneity, *Unpublished manuscript*, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=908802.

Fader, P.S., B.G.S Hardie, and K.L. Lee (2005a), RFM and CLV: Using Iso-Value Curves for Customer Base Analysis, *Journal of Marketing Research* 42 (4), 415-430.

Fader, P.S., B.G.S. Hardie, and K. Jerath (2006), Estimating CLV Using Aggregated Data: The Tuscan Lifestyles Case Revisited, *Unpublished manuscript*, http://marketing.wharton.upenn.edu/documents/research/Fader_hardie_jerath_jim_07.pdf.

Fader, P.S., B.G.S. Hardie, and K.L. Lee (2005b), "Counting Your Customers" The Easy Way: An Alternative to The Pareto/NBD Model, *Marketing Science* 24 (2), 275-284.

Fader, P.S., B.G.S. Hardie, and P.D. Berger (2004), Customer-Base Analysis With Discrete-Time Transaction Data, *Unpublished manuscript*, http://www.brucehardie.com/papers/BGBB_2004-09-25.pdf.

Ferreira, J.B., M. Vellasco, M.A. Pacheco, and C.H. Barbosa (2004), Data Mining Techniques on the Evaluation of Wireless Churn, *Proceedings of the European Sysmposium on Artificial Neural Networks* 483-488.

Forbes, T. (2007), Valuing Customers, *Database Marketing & Customer Strategy Management* 15 (1), 4-10.

Gerpott, T.J., W. Rams, and A. Schindler (2001), Customer Retention, Loyalty, and Satisfaction in the German Mobile Cellular Telecommunications Market, *Telecommunications Policy* 25 (4), 249-269.

Gjessing, H.K., O.O. Aalen, and N.L. Hjort (2003), Frailty Models Based on Lévy Processes, *Advances in Applied Probability* 35 (2), 532-550.

Glidden, D.V. (1998), Checking the Adequacy of the Gamma Frailty Model for Multivariate Failure Times, *Biostatistics Technical Report*, Technical Report No. 57, University Of California.

Gloy, B.A., J.T. Akridge, and P.V. Preckel (1997), Customer Lifetime Value: An Application in the Rural Petroleum Market, *Agribusiness* 13 (3), 335-347.

Grambsch, P.M., and T.M. Therneau (1994), .Proportional Hazards Tests and Diagnostics Based on Weighted Residuals, *Biometrika* 81 (3), 515-26.

Grambsch, P.M., T.M. Therneau, and T.R. Fleming (1995), Diagnostic Plots to Reveal Functional Form for Covariates in Multiplicative Intensity Models, *Biometrics* 51 (4), 1469-1482.

Grant, A.W.H., and L.A. Schlesinger (1995), Realize your Customers' Full Profit Potential, *Harvard Business Review* 73 (5), 59-72.

Gupta, S., and D.R. Lehmann (2003), Customers as Assets, *Journal of Interactive Marketing* 17 (1), 9-24.

Gupta, S., and D.R. Lehmann (2005), *Managing Customers as Investments: The Strategic Value of Customers in the Long Run.* New Jersey: Wharton School Publishing.

Gupta, S., and V. Zeithaml (2006), Customer Metrics and Their Impact on Financial Performance, M*arketing Science* 25 (6), 718-739.

Gupta, S., D. Hanssens, B. Hardie, W. Kahn, V. Kumar, N. Lin, N. Ravishanker, and S. Sriram (2006), Modeling Customer Lifetime Value, *Journal of Service Research* 9 (2), 139-155.

Gupta, S., D.R. Lehmann, and J.A. Stuart (2004), Valuing Customers, *Journal of Marketing Research* 41 (1), 7-18.

Gurau, C., and A. Ranchhold (2002), How to Calculate the Value of a Customer – Measuring Customer Satisfaction: A Platform for Calculating, Predicting and Increasing Customer Profitability, *Journal of Targeting, Measurement and Analysis for Marketing* 10 (3), 203-219.

Gustafsson, A., M.D. Johnson, and I. Roos (2005), The Effects of Customer Satisfaction, Relationship Commitment Dimensions, and Triggers on Customer Retention, *Journal of Marketing* 69 (4), 210-218.

Gutierrez, R.G. (2002), Parametric Frailty and Shared Frailty Survival Models, *The STATA Journal* 2 (1), 22-44.

Hadden, J., A. Tiwari, R. Roy, and D. Ruta (2005), Computer Assisted Customer Churn Management: State-Of-The-Art and Future Trends, *Computers & Operations Research* 34 (10), 2902-2917.

Haenlein, M, A.M. Kaplan, and D. Schoder (2006), Valuing the Real Option of Abandoning UnprofiTable Customers When Calculating Customer Lifetime Value, *Journal of Marketing* 70(3), 5-20.

Hansotia, B. (2004), Company Activities for Managing Customer Equity, *Journal of Database Marketing & Customer Strategy Management* 11 (4), 319-332.

Heckman, J.J., and B. Singer (1982), The Identification Problem in Econometric Models for Duration Data. In W. Hildenbrand (Ed.), *Advances in econometrics*, Cambridge: *Cambridge University Press*, 39-77.

Henderson, R., and P. Oman (1999), Effect of Frailty on Marginal Regression Estimates in Survival Analysis, *Journal of the Royal Statistical Society: Series B* 61 (2), 367-379.

Hertz-Picciotto, I., and B. Rockhill (1997), Validity and Efficiency of Approximation Methods for Tied Survival Times in Cox Regression, *Biometrics* 53 (3), 1151-1156.

Hess, K.R. (1995), Graphical Methods for Assessing Violations of the Proportional Hazards Assumption in Cox Regression, *Statistics in Medicine* 14 (15), 1707-1723.

Hoekstra, J.C., and E.K.R.E. Huizingh (1999), The Lifetime Value Concept in Customer-Based Marketing, *Journal of Market Focused Management* 3 (3/4), 257-274.

Hogan, J.E., D.R. Lehmann, M. Merino, R.K. Srivastava, J.S. Thomas, and P.C. Verhoef (2002a), Linking Customer Assets to Financial Performance, *Journal of Service Research* 5 (1), 26-38.

Hogan, J.E., K.N. Lemon, and R.T. Rust (2002b), Customer Equity Management: Charting New Directions for The Future Marketing, *Journal of Service Research* 5 (1), 4-12.

Homburg, C., M. Droll, and D. Totzek (2008), Customer Prioritization: Does It Pay Off, And How Should It Be Implemented?, *Journal of Marketing* 72 (5), 110-130.

Hosmer, D.W., and S. Lemeshow (1999), *Applied Survival Analysis: Regression Modelling of Time to Event Data*. New York: John Wiley & Sons.

Hougaard, P. (1986), A Class of Multivariate Failure Time Distributions, *Biometrika* 73 (3), 671-768.

Hougaard, P. (1986), Survival Models for Heterogeneous Populations Derived from Stable Distributions, *Biometrika* 73 (2), 387-396.

Hougaard, P. (1995), Frailty Models for Survival Data, *Lifetime Data Analysis* 1 (3), 255-273.

Hougaard, P. (2000), *Analysis of Multivariate Survival Data*. New York: Springer.

Hung, S-Y., D.C. Yen, and H-Y. Wang (2006), Applying Data Mining to Telecom Churn Management, *Expert Systems with Applications* 31 (3), 515-524.

Hwang, H., T. Jung, and E. Suh (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), 181-188.

Jackson, B.B. (1985). *Winning and Keeping Industrial Customers*. Lexington: D.C. Heath and Company.

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

Jamal, Z., and R.E. Bucklin (2006), Improving the Diagnosis and Prediction of Customer Churn: A Heterogeneous Hazard Modeling Approach, *Journal of Interactive Marketing* 20 (3-4), 16-29.

Jenkins, S. (2005), Survival Analysis, *Lecture Notes Manuscript, Unpublished manuscript*, http://www.iser.essex.ac.uk/files/teaching/stephenj/ec968/pdfs/ec968lnotesv6.pdf.

Johnson, M.D., and F. Selnes (2004), Customer Portfolio Management: Toward a Dynamic Theory of Exchange Relationships, *Journal of Marketing* 68 (2), 1-17.

Johnson, M.D., and F. Selnes (2005), Diversifying Your Customer Portfolio, *MIT Sloan Management Review* 46 (3), 10-14.

Jones, T.O., and W.E. Sasser (1995), Why Satisfied Customers Defect, *Harvard Business Review* 73 (6), 88-99.

Kalbfleisch, J.D. , and R.L. Prentice (1980), *The Statistical Analysis of Failure Time Data*. New York: John Wiley & Sons.

Kalbfleisch, J.D., and R.L. Prentice (1973), Marginal Likelihoods Based on Cox's Regression and Life Model, *Biometrika* 60 (2), 267-278.

Kaplan, E.L., and P. Meier (1958), Nonparametric Estimation from Incomplete Observations, *Journal of the American Statistical Association* 53 (282), 457-481.

Karim, M.E. (2008), Frailty Models, *Working Paper*, University Of Shaka, Bangladesh.

Keele, L.J. (2008), Proportionally Difficult: Testing for Nonproportional Hazards in Cox Models, *Working Paper*, Department Of Political Science, Ohio State University.

Kiefer, N.M. (1988), Economic Duration Data and Hazard Functions, *Journal of Economic Literature* 26 (2), 646 – 679.

Kim, H-S., and C-H. Yoon (2004), Determinants of Subscriber Churn and Customer Loyalty in the Korean Mobile Telephony Market, *Telecommunications Policy* 28 (9-10), 751-765.

Kim, Y.A., H.S. Song, and S.H. Kim (2005), Strategies for Preventing Defection Based on the Mean Time to Defection and Their Implementations on a Self-Organizing Map, *Expert Systems* 22 (5), 265-278.

Klein, J.P., and M.L. Moeschberger (1997), *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer.

Klein, J.P., and M.L. Moeschberger (2003), *Survival Analysis: Techniques for Censored and Truncated Data* (2d ed.). New York: Springer.

Kleinbaum, D.G., and M. Klein (2005), *Survival Analysis: A Self-Learning Text* (2nd Edition). New York: Springer.

Kotler, P. (1974), Marketing During Periods of Shortage, *Journal of Marketing* 38 (3), 20-29.

Kumar, V., and D. Shah (2004), Building and Sustaining ProfiTable Customer Loyalty for the 21st Century, *Journal of Retailing* 80 (4), 317-330.

Kumar, V., and M. George (2007), Measuring and Maximising Customer Equity: A Critical Analysis, *Journal of the Academy of Marketing Science* 35 (2), 157-171.

Kumar, V., G. Ramani, and T. Bohling (2004), Customer Lifetime Value Approaches and Best Practice Applications, *Journal of Interactive Marketing* 18 (3), 60-72.

Kumar, V., K.N. Lemon, and A. Parasuraman (2006), Managing Customers for Value: An Overview and Research Agenda, *Journal of Service Research* 9 (2), 87-94.

Lagakos, S.W., and D.A. Schoenfeld (1984), Properties of Proportional-Hazards Score Tests under Misspecified Regression Models, *Biometrics* 40 (4), 1037-1048.

Lee, E.T., and J.W. Wang (2003), *Statistical Methods for Survival Data Analysis* (3rd Edition). New Jersey: John Wiley & Sons.

Lee, J.H, and S.C. Park (2005), Intelligent ProfiTable Customers Segmentation System Based on Business Intelligence Tools, *Expert Systems with Applications* 29 (1), 145-152.

Lemmens, A., and C. Croux (2006), Bagging and Boosting Classification Trees to Predict Churn, *Journal of Marketing Research* 43 (2), 276-286.

Lemon, K.N., R.T. Rust, and V.A. Zeithaml (2001), What Drives Customer Equity?, *Marketing Management* 10 (1), 20-25.

Levin, N., and J. Zahavi (2001), Predicting Modeling Using Segmentation, *Journal of Interactive Marketing* 15 (2), 2-22.

Li, T., P. Lu, Z. He, and Q. Wang (2006), A Customer Retention System Based on the Customer Intelligence for a Telecom Company, *Proceedings of the Joint Conference on Information Sciences*.

Lin, D.Y., (1994) Cox Regression Analysis of Multivariate Failure Time Data, *Statistics in Medicine* 13 (21), 2233-2247.

Lu, J. (2002), Predicting Customer Churn in the Telecommunications Industry – An Application of Survival Analysis Modelling Using SAS, *Proceedings of the SAS Conference*.

Lu, J. (2003), Modeling Customer Lifetime Value Using Survival Analysis – An Application In The Telecommunications Industry, *Proceedings of the SAS Conference*.

Madden, G., S.J. Savage, and G. Coble-Neal (1999), Subscriber Churn in the Australian ISP Market, *Information Economics and Policy* 11 (2), 195-207.

Malthouse, E.C., and R.C. Blattberg (2004), Can We Predict Customer Lifetime Value?, *Journal of Interactive Marketing* 19 (1), 2-16.

Mani, D.R., J. Drew, A. Betz, and P. Datta (1999), Statistics and Data Mining Techniques for Lifetime Value Modeling, *Proceedings of the fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 94-103.

McClymont, H. and G. Jocumsen (2003), How to Implement Marketing Strategies Using Database Approaches, *Journal of Database Marketing & Customer Strategy Management* 11 (2), 135-148.

McGilchrist, C.A., and C.W. Aisbett (1991), Regression with Frailty in Survival Analysis, *Biometrics* 47 (2), 461-466.

Menezes, R. (2004), Estimation Issues on the Proportional Hazards Function Applied to Joblessness Duration Data, in Ferreira, M.A., R. Menezes e F. Catanas (eds.), *Temas em Métodos Quantitativos*, n.º 4, Lisboa: *Edições Sílabo* 131-138.

Mittal, V., M. Sarkees, and F. Murshed (2008), The Right Way to Manage Unprofitable Customers, *Harvard Business Review* 86 (4), 95-102.

Mondschein, S., and A. Musalem (2004), How Much Should You Invest in Each Customer Relationship? A Stochastic Dynamic Programming Approach, *Unpublished manuscript*, http://faculty.fuqua.duke.edu/~amusalem/bio/Musalem%20Joshi%20MktgSci%20final.pdf.

Mozer, M.C., R. Wolniewicz, D.B. Grimes, E. Johnson, and H. Kaushansky (2000a), Churn Reduction in the Wireless Industry, *Advances in Neural Information Processing Systems 12,* Cambridge: MIT Press, 935-941.

Mozer, M.C., R. Wolniewicz, D.B. Grimes, E. Johnson, and H. Kaushansky (2000b), Predicting Subscriber Dissastisfaction and Improving Retention in the Wireless Telecommunications Industry, *IEEE Transactions on Neural Networks - Special Issue on Data Mining and Knowledge Representation* 11 (3), 690-696.

Mulhern, F.J. (1999), Customer profitability Analysis: Measurement, Concentration, and Research Directions, *Journal of Interactive Marketing* 13 (1), 25-40.

Mutanen, T. (2006), Customer Churn Analysis – A Case Study, *Technical Research Centre of Finland (VTT)*, Research report no. VTT-R-01184-06.

Nasr-Bechwati, N., and A. Eshghi (2005), Customer Lifetime Value Analysis: Challenges and Words of Caution, *The Marketing Management Journal* 15 (2), 87-97.

Nath, S.V., and R.S. Behara (2003), Customer Churn Analysis in the Wireless Industry: A Data Mining Approach, *Unpublished manuscript*, http://download.oracle.com/owsf_2003/40332.pdf.

Neslin, S.A., S. Gupta, W. Kamakura, J. Lu, and C. H. Mason (2006), Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models, *Journal of Marketing Research* 43 (2), 204-211.

Ng'Andu, N.H. (1997), An Empirical Comparison of Statistical Tests for Assessing the Proportional Hazards Assumption of Cox's Model, *Statistics in Medicine* 16 (6), 611-626.

Niraj, R., M. Gupta, and C. Narasimhan (2001), Customer profitability in a Supply Chain, *Journal of Marketing* 65 (3), 1-16.

Parmar, M.K.B., and D. Machin (1995), *Survival Analysis: A Practical Approach.* Chichester: John Wiley & Sons.

Payne, A., and P. Frow (1999), Developing a Segmented Service Strategy, *Journal of Marketing Management* 15 (8), 797-818.

Payne, A., S. Holt, and P. Frow (2009), Integrating Employee, Customer and Shareholder Value Through an Enterprise Performance Model: An Opportunity for Financial Services, *The International Journal of Bank Marketing* 18 (6), 258-273.

Peppers, D., and M. Rogers (2004), *Managing Customer Relationships – A Strategic Framework.* New Jersey: John Wiley & Sons.

Peppers, D., and M. Rogers (2005), *Return on Customers: Creating Maximum Value from Your Scarcest Resource.* New York: Doubleday.

Pettersson, M. (2004), SPC with Applications to Churn Management, *Quality and Reliability Engineering International* 20 (5), 397-406.

Pfeifer, P.E., and H. Bang (2005), Non-Parametric Estimation of Mean Customer Lifetime Value, *Journal of Interactive Marketing* 19 (4), 48-66.

Pfeifer, P.E., and P.W. Farris (2004), The Elasticity of Customer Value to Retention: The Duration of a Customer Relationship, *Journal of Interactive Marketing* 18 (2), 20-31.

Pfeifer, P.E., and R.L. Carraway (2000), Modeling Customer Relationships as Markov Chains, *Journal of Interactive Marketing* 14 (2), 43-55.

Pfeifer, P.E., M.E. Haskins, and R.M. Conroy (2005), Customer Lifetime Value, Customer Profitability, and the Treatment of Acquisition Spending, *Journal of Managerial Issues* 17 (1), 11-25.

Putter, H., M. Fiocco, and R.B. Geskus (2007), Tutorial in Biostatistics: Competing Risks and Multi-State Models, *Statistics in Medicine* 26 (11), 2389-2430.

Qian, Z., W. Jiang, and K-L. Tsui (2006), Churn Detection via Customer Profile Modelling, *International Journal of Production Research* 44 (14), 2913-2933.

Ranaweera, C, and J. Prabhu (2003), On the Relative Importance of Customer Satisfaction and Trust as Determinants of Customer Retention and Positive Word of Mouth, *Journal of Targeting, Measurement and Analysis for Marketing* 12 (1), 82-90.

Ranaweera, C. (2007), Are Satisfied Long-Term Customers More Profitable? Evidence From the Telecommunication Sector, *Journal of Targeting, Measurement and Analysis for Marketing* 15 (2), 113-120.

Reichheld, F.F. (1991/92), The Truth of Customer Retention, *Journal of Retail Banking* 13 (4), 21-24.

Reichheld, F.F. (1993), Loyalty-Based Management, *Harvard Business Review* 71 (2), 64-73.

Reichheld, F.F. (1996), Learning from Customer Defections, *Harvard Business Review* 74 (2), 56-69.

Reichheld, F.F., and D. Kenny (1990), The Hidden Advantages of Customer Retention, *Journal of Retail Banking* 12 (4), 19-23.

Reichheld, F.F., and W.E. Sasser (1990), Zero Defections: Quality Comes to Services, *Harvard Business Review* 68 (5), 105-111.

Reinartz, W., and V. Kumar (2000), On the Profitability of Long-Life Customers in a Noncontractual Setting: An Empirical Investigation and Implications for Marketing, *Journal of Marketing* 64 (4), 17-35.

Reinartz, W., and V. Kumar (2002), The Mismanagement of Customer Loyalty, *Harvard Business Review* 80 (7), 86-94.

Reinartz, W., and V. Kumar (2003), The Impact of Customer Relationship Characteristics on ProfiTable Lifetime Duration, *Journal of Marketing* 67 (1), 77-99.

Reinartz, W., J.S. Thomas, and V. Kumar (2005), Balancing Acquisition and Retention Resources to Maximize Customer Profitability, *Journal of Marketing* 69 (1), 63-79.

Rosenberg, L.J., and J.A. Czepiel (1983), A Marketing Approach for Customer Retention, *The Journal of Consumer Marketing* 1 (2), 45-51.

Rosset, S., and E. Neumann (2003), Integrating Customer Value Considerations into Predictive Modelling, *Proceedings of the Third IEEE International Conference on Data Mining*.

Rosset, S., E. Neumann, U. Eick, and N. Vatnik (2003), Customer Lifetime Value Models for Decision Support, *Data Mining and Knowledge Discovery* 7 (3), 321-339.

Rosset, S., E. Neumann, U. Eick, N. Vatnik, and Y. Idan (2002), Customer Lifetime Value Modelling and Its Use for Customer Retention Planning, *Proceedings of the Eight ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 332-340.

Rust, R.T., K.N. Lemon, and V.A. Zeithaml (2004), Return on Marketing: Using Customer Equity to Focus Marketing Strategy, *Journal of Marketing* 68 (1), 109-127.

Rust, R.T., V.A. Zeithaml, and K.N. Lemon (2000), *Driving Customer Equity – How Customer Lifetime Value is Reshaping Corporate Strategy*. New York: The Free Press.

Ryals, L. (2002a), Are Your Customers Worth More than Money?, *Journal of Retailing and Consumer Services* 9 (5), 241-251.

Ryals, L. (2002b), Measuring Risk and Returns in the Customer Portfolio, *Journal of Database Marketing* 9 (3), 219-227.

Ryals, L. (2003a), Creating Profitable Customers Through the Magic of Data Mining, *Journal of Targeting, Measurement and Analysis for Marketing* 11 (4), 343-349.

Ryals, L. (2003b), Making Customers Pay: Measuring and Managing Customer Risk and Returns, *Journal of Strategic Marketing* 11 (3), 165-175.

Ryals, L. (2005), Making Customer Relationship Management Work: The Measurement and ProfiTable Management of Customer Relationships, *Journal of Marketing* 69 (4), 252-261.

Ryals, L. (2006), Profitable Relationships With Key Customers: How Suppliers Manage Pricing and Customer Risk, *Journal of Strategic Marketing* 14 (2), 101-113.

Ryals, L., and S. Knox (2007), Measuring and Managing Customer Relationship Risk in Business Markets, *Industrial Marketing Management* 36 (6),1-11.

Schemper, M. (1992), Cox Analysis of Survival Data with Non-Proportional Hazard Functions, *The Statistician* 41 (4), 455-465.

Schmittlein, D.C., and R.A. Peterson (1994), Customer Base Analysis: An Industrial Purchase Process Application, *Marketing Science* 13 (1), 41-67.

Schmittlein, D.C., D.G. Morrison, and R. Colombo (1987), Counting Your Customers: Who Are They And What Will They Do Next?, *Management Science* 33 (1), 1-24.

Schwarz, G., (1978), Estimating the Dimension of a Model, *The Annals of Statistics* 6 (2), 461-464.

Schweidel, D.A., P.S. Fader, and E.T. Bradlow (2006), Modeling Retention Within and Across Cohorts, *Unpublished manuscript*, http://marketing.wharton.upenn.edu/ideas/pdf/Fader/PapersNew/cohort_retention_1005.pdf.

Seo, D.B., C. Ranganathan, and Y. Babad (2007), Two-Level Model of Customer Retention in the US Mobile Telecommunications Service Market, *Telecommunications Policy* 32 (3-4)1-15.

Sevin, Charles H. (1965), *Marketing Productivity Analysis*. St. Louis: McGraw-Hill.

Shah, D., R.T. Rust, A. Parasuraman, R. Staelin, and G.S. Day (2006), The Path to Customer Centricity, *Journal of Service Research* 9 (2), 113-124.

Shapiro, B.P., V.K. Rangan, R.T. Moriarty, and E.B. Ross (1987), Manage Customers for Profits (Not Just Sales), *Harvard Business Review* 65 (5), 101-108.

Shih, Y., and C. Liu (2003), A Method For Customer Lifetime Value Ranking – Combining the Analytic Hierarchy Process and Clustering Analysis, *Journal of Database Marketing & Customer Strategy Management* 11 (2), 159-172.

Singer, J.D., and J.B. Willett (2003), *Applied Longitudinal Data Analysis – Modeling Change and Event Occurrence.* New York: Oxford University Press.

Singh, S.S. (2003), Customer Lifetime Value Analysis, *Unpublished Doctoral Dissertation*, Northwestern University.

Song, H.S., J.K. Kim, Y.B. Cho, and S.H. Kim (2004), A Personalized Defection Detection and Prevention Procedure Based on the Self-Organizing Map and Association Rule Mining: Applied to Online Game Site, *Artificial Intelligence Review* 21 (2), 161-184.

Stahl, H.K., K. Matzler, and H.H. Hinterhuber (2003), Linking Customer Lifetime Value with Shareholder Value, *Industrial Marketing Management* 32 (4), 267-279.

Struthers, C.A., and J.D. Kalbfleisch (1984), Misspecified Proportional Hazard Models, *Biometrika* 73 (2), 363-369.

Tableman, M., and J.S. Kim (2004), *Survival Analysis with S: Analysis of Time-To-Event Data.* Boca Raton: Chapman & Hall/CRC.

Thakur, R., and J.H. Summey (2005), Filtering Profitable from Not-So-Profitable Customers Using Customer Relationship Management (CRM) Technology, *The Marketing Management Journal* 15 (2), 43-54.

Therneau, T.M., and P.M. Grambsch (2000), *Modeling Survival Data: Extending the Cox Model.* New York: Springer.

Therneau, T.M., P.M. Grambsch, and T.R. Fleming (1990), Martingale-Based Residuals for Survival Models, *Biometrika* 77 (1), 147-160.

Therneau, T.M., P.M. Grambsch, and V.S. Pankratz (2000), Penalized Survival Models and Frailty, *Technical Report No. 66*, Mayo Foundation.

Thomas, J.S., W. Reinartz, and V. Kumar (2004), Getting the Most Out of All Your Customers, *Harvard Business Review* 82 (7/8), 116-123.

Trubik, E., and M. Smith (2000), Developing a Model of Customer Defection in the Australian Banking Industry, *Managerial Auditing Journal* 15 (5), 199-215.

Turnbull, P.W. (1990), A Review of Portfolio Planning Models for Industrial Marketing and Purchasing Management, *European Journal of Marketing* 24 (3), 7-22.

van Amelsfort, L.K. (2006), Influencing Uncertainty in Customer Equity by Changing Customer Portfolio Characteristics, *Unpublished Master Dissertation in Mathematical Economics and Econometric Models Thesis*, Universiteit van Tilburg

Van den Poel, D., and B. Larivière (2004), Customer Attrition Analysis for Financial Services Using Proportional Hazard Models, *European Journal of Operational Research* 157 (1), 196-217.

Vaupel, J.W., K. Manton, and E. Stallard (1979), The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality, *Demography* 16 (3), 439-454.

Venkatesan, R., and V. Kumar (2004), A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy, *Journal of Marketing* 68 (4), 106-125.

Villanueva, J. (2003), Acquisition channels and price discrimination in a customer equity framework, *Unpublished Doctoral Dissertation*, University of California, Los Angeles.

Villanueva, J., and D.M. Hanssens (2007), Customer Equity: Measurement, Management and Research Opportunities, *Foundations and Trends in Marketing* 1 (1), 1-95.

von Wangenheim, F. (2006), Lifetime Value Prediction at Early Relationship Stages, *Working Paper*, University of Dortmund.

von Wangenheim, F., and P. Lentz (2005), Understanding Your Customer Portfolio: A Simple Approach to Customer Segmentation According to Lifecycle Dynamics, *Working paper*, University of Dortmund.

Voss, K.E. (1998), The Expected Value of a Customer: A Model and Applications, *Unpublished Doctoral Dissertation*, Washington State University.

Wayland, R.E., and P.M. Cole (1994), Turn Customer Service into Customer Profitability, *Management Review* 83 (7), 22-24.

Wei, C-P, and I.T. Chiu (2002), Turning Telecommunications Call Details to Churn Prediction: A Data Mining Approach, *Expert Systems with Applications* 23 (2), 103-112.

Weinstein, A. (2002), Customer-Specific Strategies – Customer Retention: A Usage Segmentation and Customer Value Approach, *Journal of Targeting, Measurement and Analysis for Marketing* 10 (3), 259-268.

Wienke, A. (2003), Frailty Models, *MPIDR Working Paper WP 2003-032*.

Wienke, A., K. Arbeev, I. Locatelli, and A.I. Yashin (2003), A Simulation Study of Different Correlated Frailty Models and Estimation Strategies, *Working Paper No. 18*, Max Planck Institute for Demographic Research.

Woo, J.Y., S.M. Bae, and S.C. Park (2005), Visualization Method for Customer Targeting Using Customer Map, *Expert Systems with Applications* 28 (4), 763-772.

Woo, K., and H.K.Y. Fock (2004), Retaining and Divesting Customers: An Exploratory Study of Right Customers, “At-Risk” Right Customers, and Wrong Customers, *The Journal of Services Marketing* 18 (2/3), 187-197.

Wyner, G.A. (1996), Customer Valuation: Linking Behavior and Economics, *Marketing Research* 8 (2), 36-38.

Yan, L., D.J. Millar, M.C. Mozer, and R. Wolniewicz (2001), Improving Prediction of Customer Behaviour in Nonstationary Environments, *Proceedings of the International Joint Conference on Neural Networks (vol 3)* 2258-2263.

Yang, A.X. (2005), Using Lifetime Value to Gain Long-Term Profitability, *Journal of Database Marketing & Customer Strategy Management* 12 (2), 142-152.

Yashin, A.I., I.A. Iachine, A.Z. Begun, and J.W. Vaupel (2001), Hidden Frailty: Myths and Reality, *Research Report 34*, Department of Statistics and Demography, CDU-Cdense University.

Zdravkovic, S., A. Wienke, N.L. Pedersen, M.E. Marenberg, A.I. Yashin, and U. de Faire (2004), Genetic Influences on CHD-Death and the Impact of Known Risk Factors: Comparison of Two Frailty Models, *Behaviour Genetics* 34 (6), 585-592.

Zeithaml, V.A., R.T. Rust, and K.N. Lemon (2001), The Customer Pyramid: Creating and Serving Profitable Customers, *California Management Review* 43 (4), 118-142.

Zhang, Y., J. Qi, H. Shu, and Y. Li (2006), Case Study on CRM: Detecting Likely Churners with Limited Information of Fixed-Line Subscriber, *Proceedings of the International Conference on Service Systems and Service Management (vol 2)* 1495-1500.

Zorn, C.W. (2000), Modeling Duration Dependence, *Political Analysis* 8 (4), 367-380.

APPENDICES