The Duration of the Customer Relationship with Fixed Telecommunications Service Providers in Portugal using Survival Analysis

Sofia PORTELA
Department of Quantitative Methods, ISCTE Business School
Lisbon, Portugal
smlportela@gmail.com

Rui MENEZES
Department of Quantitative Methods, ISCTE Business School
Lisbon, Portugal
rui.menezes@iscte.pt

ABSTRACT

Customer churn is the customer’s decision to terminate the relationship with a provider. This decision can be very onerous to the business financial performance. As such, an a priori knowledge about the probability (risk) of a given customer to cancel the relationship with the firm in the next period is a valuable tool that allows firms to take preventive measures to avoid the defection of potentially profitable customers. This study aims to understand and predict customer lifetime in a contractual setting in order to improve the practice of customer portfolio management. A duration model is developed to understand and predict the residential customer churn in the fixed telecommunications industry in Portugal. The model is developed by using large-scale data from an internal database of a Portuguese company which presents bundled offers of ADSL, fixed line telephone, pay-TV and home-video. The model is 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..

KEY WORDS: Survival analysis, duration models, customer churn, customer retention, customer management.

1 INTRODUCTION

Customer churn, i.e., the customer’s decision to terminate the relationship with a provider, is a major concern for fixed telecommunications firms in Portugal. In fact, the considerable increase of business competition in the Portuguese fixed telecommunications industry over the last decades has given rise to a phenomenon of customer switching behaviour, and, thus, high customer churn rates, which has serious consequences for the financial performance of the firms and, therefore, for the economy. Several researchers have mentioned that customer churn is the main reason of profitability losses in the telecommunications industry, due to losses on current and potential revenues, marketing costs, and brand image (e.g., [1, 26, 33]).

As a consequence of this steady market growth, firms have been focused on customer acquisition and neglected customer retention. Nevertheless, the fixed telecommunications market is becoming saturated in Portugal and, as a consequence, the pool of “available customers” is limited and firms need to change their strategy from customer acquisition to customer retention [14, 16].

Customer retention became a buzzword in the 1990s, mainly due to the work of [6], who firstly provided evidence about the advantages of customer retention. Even though their results are not consensual (see, for example, [6, 9, 10, 12, 28]) they definitively caused a change in the marketing theory.

Despite the large amount of research done on customer churn in mobile telecommunications, there are few studies applied to the fixed telecommunications industry and none applied to firms that provide bundled offers of fixed telecommunications services. Moreover, this issue has never been studied in Portugal. So, this study aims to develop a model of the residential customer churn in the fixed telecommunications industry in Portugal. Specifically, this study intends to estimate the probability of a given active customer cancels his/her relationship with the firm in the next period. Some of the specific areas where this model can help customer management are: (i) a priori knowledge about the probability (risk) of a given customer to cancel the relationship with the firm in the next period and, in this way, firms can take preventive measures to avoid the defection of potentially profitable customers, (ii) customer selection to retention programs; (iii) marketing resource allocation across customers; and (iv) computation of customer lifetime value..

2 METHODOLOGY

A continuous-time duration model will be used to understand the residential customer churn in the fixed telecommunications industry (contractual settings) in Portugal. Let $T$ be a continuous-time non-negative random variable, which represents the survival time in days. The distribution of $T$ is described by four equivalent functions: (i) survival function; (ii) density function; (iii) hazard function; and (iv) cumulative hazard function. The cumulative density function of $T$ is expressed as

$$F(t) = P(T \leq t) = \int_0^t f(u) \, du \quad (1).$$

The survival function is the probability of a customer to last beyond time $t$ [8], that is

$$S(t) = P(T > t) = 1 - F(t) \quad (2).$$

---

1 By fixed telecommunications industry we mean firms that provide fixed-line telephone, internet, and pay-TV.
The density function is given by

\[ f(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t} = \frac{dF(t)}{dt} \]  \hspace{1cm} (3).

The hazard function is the instantaneous potential per unit time for the event occurrence, given that the customer has survived up to time \( t \). The hazard function is given by

\[ h(t) = \lim_{\Delta t \to 0} \frac{P(t < T < t + \Delta t | T > t)}{\Delta t} \cdot \frac{f(t)}{S(t)} \]  \hspace{1cm} (4).

The cumulative hazard function is the total risk accumulated up to time \( t \) that is

\[ H(t) = \int_{0}^{t} h(x) \, dx = -\int_{0}^{t} \frac{1}{S(x)} \left[ \frac{d}{dx} S(x) \right] \, dx \]  \hspace{1cm} (5).

Duration models can accommodate both the proportional hazards (PH) and the accelerated failure time (AFT) forms. PH models assume that the hazard rates of any two customers are proportional over time (which means that the hazard ratio is constant over time) (PH assumption) [31], and, as such, the effect of any covariate in the hazard function is constant over time [2, 5]. That is,

\[ h_i(t|X_j) = HR \times h_j(t|X_j) \cdot HR \geq 0 \]  \hspace{1cm} (6).

PH models are expressed as

\[ h(t|X) = h_0(t) \cdot \lambda \cdot h_0(t) \geq 0, \lambda > 0 \]  \hspace{1cm} (7).

where \( h_0(t) \) is the baseline hazard function and \( \lambda = \exp(\beta' X) \).

AFT models are linear models of \( \ln(T) \), that is

\[ \ln(T) = \beta' X + \sigma \varepsilon \]  \hspace{1cm} (8),

where \( \varepsilon \) is a random disturbance with a fixed variance and \( \sigma \) is a scale parameter that controls the variance of \( \varepsilon \). This model assumes that \( \varepsilon \) is independent of \( X \) and that \( \varepsilon_i \) is independent of \( \varepsilon_j, i \neq j \) [2]. In order to estimate the AFT model by maximum likelihood, the probability distribution of \( \varepsilon \) must be specified [2, 7]. Four distributions are usually used for \( \varepsilon \): normal, logistic, extreme value, and log-gamma [2]. The distribution of \( T \) depends on the distribution of \( \varepsilon \), as presented in [2, p. 374].

AFT models assume that there is a constant non-negative acceleration factor that stretches out or shrinks survival times [8], that is

\[ S_i(t) = S_0(\psi t) \cdot t \geq 0, \psi \geq 0 \]  \hspace{1cm} (9),

where \( S_i(t) \) is the probability that the contract \( i \) survives beyond time \( t \). \( S_0 \) is the baseline survival (the value of all covariates are equal to zero), and \( \psi = \exp(-\beta' X) \) is the acceleration factor. It is expected that the event of interest occurs sooner for contracts with \( \psi > 1 \) and later for contracts with \( \psi < 1 \) [7].

Duration models can be divided into three types: nonparametric models, semi-parametric models and parametric models.

Nonparametric models do not make any assumption about the shape of the relevant functions and they do not include any covariate [8]. As such, the hazard function is estimated only on the basis of empirical data of survival time and on the customer status.

As regards to semi-parametric models, the baseline hazard is unknown and left-unparameterized, which means that no assumption about the shape of the baseline hazard function is made [5, 7, 22] and thus the baseline hazard can assume any shape. Semi-parametric models are PH models. Semi-parametric models are estimated by the partial likelihood method. The well-known Cox Proportional Hazards Model is a semi-parametric model and its hazard function is given by

\[ h(t|X) = h_0(t)\exp(\beta' X) \cdot h_0(t) \geq 0 \]  \hspace{1cm} (10).

Lastly, parametric models assume that the data distribution is known. In duration analysis time \( T \) can follow several known distributions, such as exponential, Weibull, Gompertz, log-normal, log-logistic and gamma distribution. Exponential and Weibull models can be parameterized as PH or AFT models; Gompertz is a PH model; and the others are AFT models. Parametric models are estimated by maximum likelihood.

The hazard function of the exponential distribution is given by

\[ \lambda = \exp(\beta' X) \]  \hspace{1cm} in the PH form and by \( \lambda = \exp(-\beta' X) \) in the AFT form [7], which means that the hazard function is constant over time [2, 8] and it does not reflect duration dependence. For this reason, the exponential distribution is often called memoryless [8, 22]. Duration dependence occurs when the hazard rate varies with the actual duration of the customer relationship. All the other distributions mentioned above account for this effect.

The hazard function of the Weibull distribution, in the PH form, is given by

\[ h(t|X) = \lambda p(t)^{p-1} \cdot \lambda > 0, p > 0 \]  \hspace{1cm} (11),

where \( \lambda = \exp(\beta' X) \), and in the AFT form, it is expressed as

\[ h(t|X) = \lambda p(\lambda t)^{p-1} \cdot \lambda > 0, p > 0 \]  \hspace{1cm} (12),

where \( \lambda = \exp(-\beta' X) \). 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 [18, 20, 21].

The hazard function of the Gompertz distribution is given by

\[ h(t|X) = \lambda \exp(\gamma t), \lambda > 0, \gamma > 0 \tag{13} \]

where \( \lambda = \exp(\beta' X) \). When \( \gamma > 0 \), the hazard function increases over time; when \( \gamma < 0 \), the hazard function falls with time; and when \( \gamma = 0 \) the hazard is flat with time and the model reduces to an exponential model [5, 22].

The hazard function of the log-normal distribution is non-monotonic and is given by

\[ h(t|X) = \frac{1}{\sigma^{2}/2\pi t^{-1}} \exp \left[ -\frac{1}{2} \left( \frac{\ln(t) - \mu}{\sigma} \right)^2 \right] \tag{14} \]

where \( \lambda = \exp(-\beta' X) \) and \( \Phi \) is the cumulative distribution function for the standard normal distribution. The hazard of the log-normal model is hump-shaped [20], which results from the log-normal distribution function for the standard normal distribution. The log-logistic distribution produces a hazard function that can be non-monotonic and unimodal. The hazard function is given by

\[ h(t|X) = \frac{\lambda p(\lambda t)^{p-1}}{1+(\lambda t)^p}, \lambda > 0, p > 0 \tag{15} \]

where \( \lambda = \exp(-\beta' X) \). When \( p > 1 \) the hazard first increases from the origin until reaches a maximum at time \( t = (p-1)^{1/p}/\lambda \) and then falls to zero as \( t \to \infty \) (similar behavior 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 starts at \( \lambda \) and then fall monotonically [18].

The generalized gamma distribution is the most flexible parameterization [21], because it allows for several possible shapes of the hazard function [5, 7]. The density function of the generalized gamma distribution is given by

\[ f(t|X) = \frac{\lambda p(\lambda t)^{pk-1} \exp(-\lambda t^p)}{\Gamma(k)} \tag{16} \]

where \( \Gamma(k) \) is the well-known gamma function.

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 gamma distribution [5, 18]. 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) [7, 20].

The duration models mentioned above assume that the covariates included in the models incorporate all the differences between the individuals. But omitted covariates may probably exist [2, 3, 15] and, in that case, their effects (usually called unobserved heterogeneity or frailties) may be included in the model [3], because when these effects are important but omitted in the model, the following consequences may happen: (i) the model will over-estimate the degree of negative duration dependence in the hazard function; (ii) the estimates of the model will be biased; and (iii) the estimates of the coefficients of the covariates in the model are not constant, but decline over time [2, 3].

The form of the unshared frailty model is

\[ h(t_j | X_j, \alpha_j) = \alpha_j h(t_j | X_j) \tag{17} \]

where \( \alpha_j \) is the frailty, i.e., individual unobserved specific effect. Frailties are latent variables that have a multiplicative effect on the hazard function and it is assumed that frailties are random positive values with mean 1 (assumed for purposes of model identification) and finite variance \( \Theta \) [3, 13, 21]. It is assumed that the omitted covariates are independent of the time \( t \), of the covariates included in the model [3] and of any censoring [15]. Frailties are not directly estimated from the data but its variance \( \Theta \) is [7, 13]. Individuals with above-average values of \( \alpha \) fail fast and individuals with below-average values of \( \alpha \) fail slowly [13, 15].

Frailty models are mixture models because the frailty \( \alpha \) has to be integrated out [3, 5, 7, 13] in order to obtain the population (or unconditional) survival function, and as such, a theoretical distribution with probability density function \( g(\alpha) \) must be specified for \( \alpha \). The most common used distribution for \( \alpha \) is the gamma distribution [3, 5, 15]. The population (or unconditional) survival function is defined as

\[ S_\theta(t | X) = \int_0^{\infty} [S(t | X)]^\alpha g(\alpha) d\alpha \tag{18} \]

where \( S_\theta \) is the population (or unconditional) survival function, that is the survival function that represents the average population and \( S \) is the individual (or conditional) survival function. The population hazard function is expressed as

\[ h_\theta(t | X) = h(t | X) E(\alpha | T > t) \tag{19} \]

which means that the population hazard function is the average hazard over the survival individuals at any given time.
3 DATA

This study is based on a dataset obtained from a Portuguese fixed telecommunications firm which presents bundled offers of ADSL, fixed line telephone, pay-TV and home-video. The database includes a random sample of 830 residential customers who completed a questionnaire about customer satisfaction. The available data contains a large number of covariates, which include customer’s basic information, demographics, churn flag, customer historical information about usage, billing, subscription, credit, and other.

4 RESULTS ANS DISCUSSION

A Cox model was estimated in order to test the PH assumption based on Schoenfeld residuals. We found statistical evidence that the PH assumption does not hold ($p = 0.004$); so, AFT models will be used instead. As all AFT models are parametric models, the data distribution has to be postulated in advance. In order to decide which parametric model is more appropriate for our data, we adopted two statistical strategies as suggested by [7]. These methods consist on (i) the estimation of a generalized gamma model and testing its free parameters and (ii) comparison of the AIC for each model. It seems that the model that best fit the data is the log-logistic.

A log-logistic model with gamma-distributed frailty was estimated in order to test for unobserved individual heterogeneity. There is statistical evidence that this effect is presented (H$_0$: $\theta = 0$; $p = 0.077$), and thus, it has to be included in the model, since it improves the results. The final model is presented in table 1.

<table>
<thead>
<tr>
<th></th>
<th>Log-logistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (0 – female; 1 – male)</td>
<td>.432 ***</td>
</tr>
<tr>
<td>Total dunning (total number of overdue bills since the beginning of the contract)</td>
<td>-.439 ***</td>
</tr>
<tr>
<td>Overall revenues (total revenues from the customer since the beginning of the contract)</td>
<td>.001 ***</td>
</tr>
<tr>
<td>Current debts</td>
<td>.008 ***</td>
</tr>
<tr>
<td>Gender $\times$ Overall revenues</td>
<td>-.000 ***</td>
</tr>
<tr>
<td>Value of off-peak calls</td>
<td>.263 **</td>
</tr>
<tr>
<td>Monthly average telephone revenues</td>
<td>-.039 ***</td>
</tr>
<tr>
<td>Monthly average internet revenues</td>
<td>-.019 ***</td>
</tr>
<tr>
<td>Constant</td>
<td>6.243 ***</td>
</tr>
<tr>
<td>Ln sigma</td>
<td>-2.466 ***</td>
</tr>
<tr>
<td>Ln theta</td>
<td>.522</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-49.248</td>
</tr>
<tr>
<td>AIC</td>
<td>120.496</td>
</tr>
</tbody>
</table>

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

Our results show that overall revenues positively affect the survival time, which is consistent with the results of [17]. Zhang et al. [33] also found that the overall revenues from the last 6 months affects the survival time. The results of the present study also appear to indicate that survival time increases as the monthly average of off-peak calls increase. 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. Ahn et al. [1] did not find any relationship between the value of current debts and survival time.

On the other hand, it seems that the total number of overdue bills (since ever), the monthly average of customer spending both on fixed-telephone and internet negatively affect the survival time. Even though some authors have found similar evidence about the monthly revenues (e.g., [1, 4, 25]), other found the opposite (e.g., [19]). This indicates that customers are very sensitive to pricing.

The results of the present study also indicate that the survival time for males is larger than the one for females, which is consistent with [1, 29], but contradicts [19, 25]. Furthermore, it seems that the effect of overall revenues on survival time for females is larger than for males.

Contrary to our prior expectations, customer satisfaction is not a significant covariate, which suggests that customer satisfaction in this context is not a reason for customer churn. A possible explanation for 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 literature about satisfaction (e.g., [4, 11]). Kim and Yoon [19] found that whereas some types of satisfaction positively affects the survival time, other do not have any influence. Van den Poel and Larivière [32] show that some studies did not also find any influence of satisfaction on survival time.

Moreover, results also appear to indicate that both the telephone and internet usage do not influence customer churn, which contrasts with the findings of [1], who found a positive relationship between usage and survival time. We also provide evidence that the payment method and the number of invoices in debt do not influence the customer churn. Nevertheless, Zhang et al. [33] show that the payment method affects the customer churn.

Lastly, the results suggest that the customer retention rate is neither constant over time (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, as mentioned above.

The hazard function of this model is presented in Figures 1.

As can be seen from the analysis of the population hazard curve, there is duration dependence. In fact, the probability that a given mean customer cancels his relationship with the service provider increases as the customer lifetime increases (for relationships with less than approximately 3.5 years), and then decreases. Different studies have also obtained duration dependence (e.g., [19, 23, 24, 29, 33]).
5 CONCLUSIONS

This study sheds new light on the crucial issue of customer churn in the fixed telecommunications industry in Portugal. Considering that it is crucial to prevent the churn of profitable customers in order to ensure the financial performance of these firms, the results of this study are very valuable mainly when complemented with an analysis of the CLV for each individual. These results have a number of managerial implications. Firstly, firms cannot make decisions about customer management based on the average churn rates. Secondly, 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.

ACKNOWLEDGEMENTS

We wish to thank the firm that provided their internal customer database, without which this study would be not possible, and who strongly believe on the importance of the expected results to the improvement of its customer portfolio management. We also thank the Fundação para a Ciência e Tecnologia for its financial support (Project reference: PTDC/GES/73418/2006).

REFERENCES

