Customer survival time in subscription-based businesses (case of Internet service providers)

Z. Mohammed, J. S. Maritz & D. Kotze
Department of Statistics, University of the Western Cape, Cape Town, South Africa

Abstract

In today's subscription-based industries of which telecommunication businesses, Internet services, medical aid, and the insurance industry are examples, management activities should be customer-centred and performance should be measured based on the firm's customer value.

In this study we investigate customer survival time in the Internet service providers industry. From the database of an Internet service provider, a random sample of 30 000 customers was selected. Factors potentially associated with churn in the ISP industry are investigated in this study. There are mainly two groups: demographic variables (gender, age, language and number of children) and usage related variables (the IT background of the customer and customer segment). Factors are examined individually. Nonparametric survival analysis techniques are used to analyse the data.

Demographics and usage related factors have shown potential in determining the survival time of the customer and the event of churn. In particular, age and number of children and the customer usage segment appeared to be extremely important. The customers with age less than 26 years have higher risk of churn, while customers who use Internet for business purposes appeared to be more loyal. A special retention strategy should be formulated for these groups.
Keywords: subscription-based business, Internet service provider, customer survival time, churn, survival analysis, survival function, hazard function.

1 Introduction

In today's subscription-based industries of which telecommunication businesses, Internet services, medical aid, and the insurance industry are examples,

management activities should be customer-centred and performance should be measured based on the firm's customer value [1–5].

Customer asset metrics – customer lifetime value (LTV) – depends on assumptions about the future stream of income of a customer, the appropriate allocation of costs to customers, the discount factor, and expected lifetime of a customer. Considering the Internet service providers (ISPs) industry, the general objective of this study is to analyse customer lifetime in this industry. We believe that customer lifetime is the most important component that requires modelling of the expected future revenue of the customer (customer lifetime value). Lu [6] emphasized that the customer survival curve and the customer monthly margin are the most important components in modelling LTV in the telecommunication industries. In addition to its use in modelling the analysis of customer survival time will be used to set appropriate retention strategies. In studying customer lifetime we use survival analysis in time-to-event data analysis [7–11]

The terms customer lifetime and customer survival time are used interchangeably, and are defined as the time from subscription to cancellation of service. The following section briefly discusses churn: its definition, types, and factors that influence it.

2 Background to churn

In business context, various definitions have been used for the term churn. It is defined to be the proportion of subscribers that leave the service out of the total subscribers at a given time period [12] or the number of customers that leave the service provider in a specific time period. A high churn rate implies high customer disloyalty. Churn is also taken to mean the act of cancelling the service or switching service providers. Two types of churn exist: voluntary and involuntary churn. The voluntary churn is initiated by the subscriber as dissatisfaction with his/her business relation to the firm. Some of the factors that influence subscriber decision to churn in the ISP industry are pricing, customer service, service, content, security and added-services as well as contract conditions. In involuntary churn the disconnection is initiated by either the service provider (as a result of default in payment and bad debt) or the customer due to unavoidable circumstance (such as subscriber death or movement out of service area). This study deals with voluntary churn where the cause of churn is widely believed to be manageable.

Factors potentially associated with churn in the ISP industry are investigated in this study. There are mainly two groups: demographic variables (gender, age, language and number of children) and usage related variables (the IT background of the customer, the use of WiFi software that speeds up the Internet and customer segment). This is a preliminary study in which the factors are examined individually. The interactions of the factors will be studied in the next phase of the analysis.

3   Methodology

From the database of an Internet service provider, a simple random sample of size 30 000 was selected. Different customers could have different entry times. Time from subscription to churn (or to date of data extraction for censored cases) was recorded in months. The values of several covariates were also recorded for every customer.

We used the Nelson–Aalen method to estimate the cumulative hazard (integrated hazard) [13, 14]. The Kaplan–Meier technique [15] was used to plot the survival curves for the different levels of each covariate (for example, for the age covariate we grouped customer into three age groups and then plotted the survival curve for each group). Equality of the survival curves of different levels for each covariate was tested by the logrank test. Both STATA and SPSS software were used to analyse the data.

4   Results

Table 1 shows the cumulative hazard of customer churn over time for different age groups. The younger customers have shown a very high risk of churn. As the age of the customer increases, the chances of churn decreases. The risk of churn is more than two times higher for the age group of a subscriber with age less than 26 years than for the subscriber with age more than 40 years.

Table 1:   Cumulative hazard of customer churn for different age groups.

| | Age group | | |
Time in months	Less than 26 years	26-40 years	More than 40 years
1	0.0102	0.0082	0.0053
13	0.0713	0.0541	0.0354
25	0.1623	0.1049	0.0677
37	0.3463	0.1841	0.1144
49	0.5497	0.2625	0.1639
61	0.6217	0.3341	0.2254
73	0.6973	0.4125	0.2962
85	0.7888	0.4977	0.3430
97	0.8211	0.5102	0.4057

In figure 1 the estimated survival curves for customers grouped according to age group are presented; they are shown with 95% confidence bands. The young customers have less survival chance while the older customers seem to be more loyal. The age factor has shown a high potential in differentiating the survival time of the subscriber. No overlaps in the survival confidence limits were seen except after a long time (that is after 93 months). The p-value of the test of
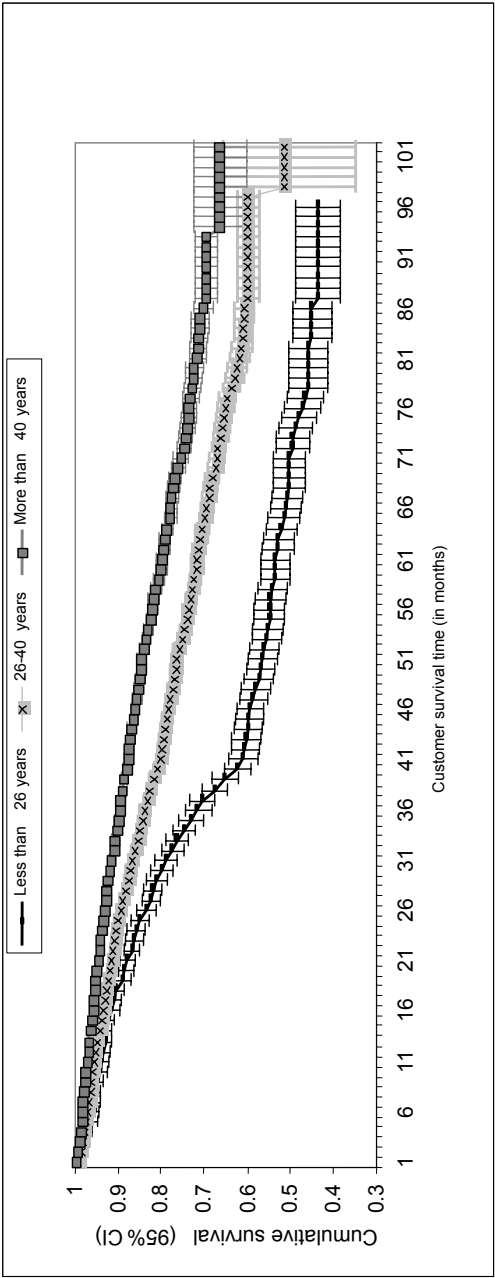
Figure 1: The survival curves for customers grouped according to age group.
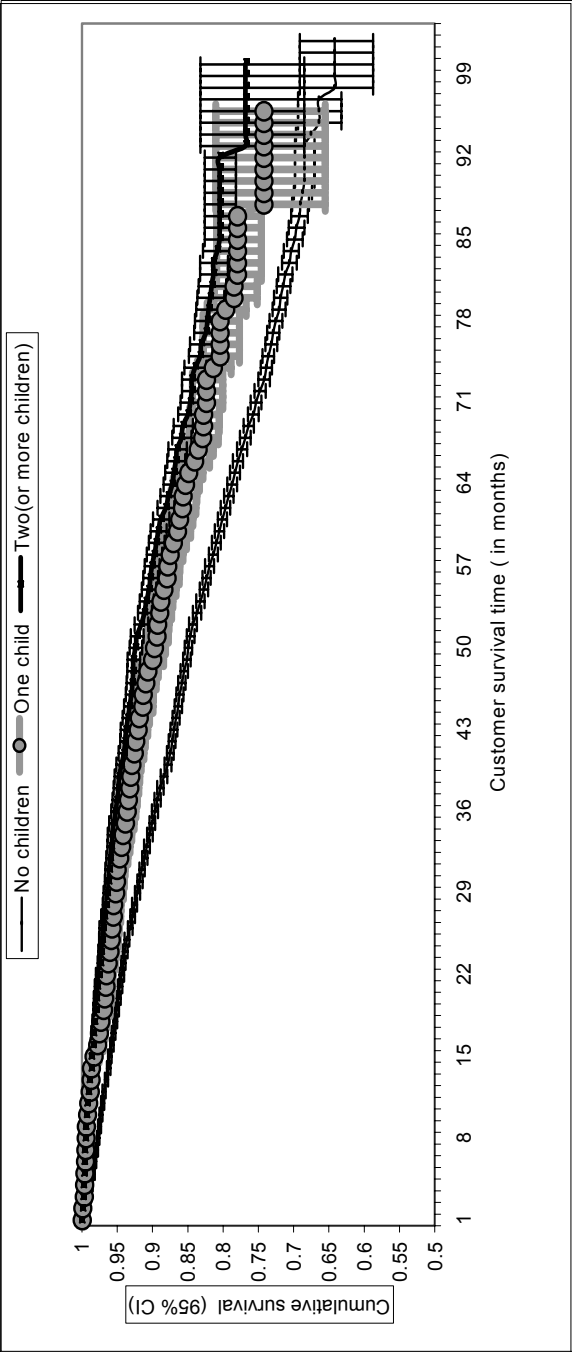
Figure 2:   The survival curves for customers grouped according to number of children.
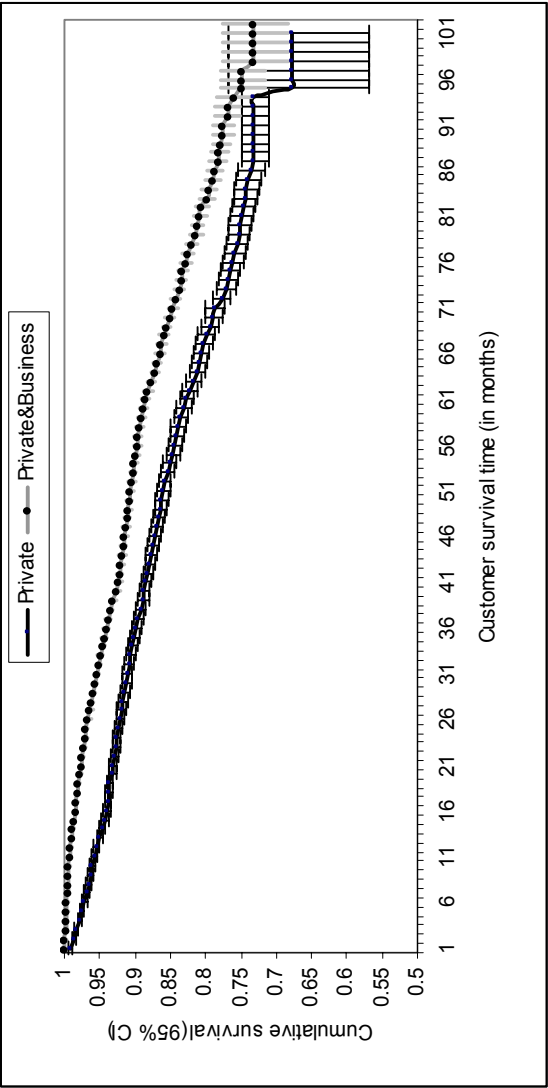
Figure 3:   Survival curves for customers grouped according to usage purpose.

equality of the three survival curves is highly significant (p = 0.000). The age group "Less than 26 years" is the only group where more than 50% of them cancelled the service before the observation was terminated; it is the only group where the median survival time of 72 months exists.

In figure 2 the estimated survival curves for customers grouped according to number of children are presented; they are shown with 95% confidence bands. The graphs indicate that families with children increase customer survival chance (loyalty). The group of customers who have two (or more) children have the best survival chance. The p-value for test of equality of the three survival curves is highly significant (p = 0.000). However, the figure shows an overlap in cumulative survival confidence limits for the group of customers who have one child and the group of customers with two (or more children). The confidence limits for the childless subscribers show a significant difference from the other two groups; most of the time no overlaps occur in the confidence limits.

In figure 3 the estimated survival curves for customers grouped according to usage purposes are presented; they are shown with 95% confidence bands. Those customers who use Internet for both private and business purposes have a better survival chance than those who use it only for private purposes. The survival curve confidence limits have shown a clear difference as there were no overlaps, except at very few points (after 92 months). The logrank test p-value for the test of equality of the two survival curves was highly significant (p = 0.000).

5   Summary, discussion and conclusion

In this study we investigated the issue of customer survival time/churn in ISP industry using two sets of factors: demographic factors (gender, language, age and number of children) and usage factors (IT background of the user, WiFi usage and customer segment). In general all factors are significantly important in determining the survival time of the customer and the event of churn.

The age factor has shown a great impact in differentiating among customers' survival time. Young customers have higher chance of churn. In our sample, among the customers of age less than 26 years old 68.14% cancelled the service. The young age group customers seem to be maybe, more enthusiastic to move around looking for better service. A special retention strategy should be formulated to focus on this group. Concerning the language factor, customers who use English as their Internet language have shown better survival chance; this could be related to the fact that they benefit the content of the Internet more than others – this is a usage related issue. The results have shown also that the number of children is a significant differentiating factor. Having children showed a positive impact on prolonging customer survival time and reducing the hazard of churn. This result could be due to the following facts: Firstly, having a family (children or wife/husband) may force you to stay with your original service provider although you may not be happy with the service because other members like it for one reason or another. Secondly, existence of other family members may increase the usage and then the importance of the Internet connection to the family. To avoid misleading interpretation, it is important to say that the fact that

number of children is an important factor may because of its high correlation with the age. The higher survival chance for customers who have children could be because they tend to be older than those with less or no children. With regard to gender, the result showed that male subscribers are more likely to be loyal.

The different levels of each usage factor – we called them usage factors because we believe that they are related to the usage – are significantly different. The existence of IT background for the customer, use of WiFi, and the use of Internet for business purposes result in a higher survival chance and less hazard of churn.

In conclusion, this paper showed the importance of the demographic and usage factors in understanding customer survival and churn in the ISP industry. Managers should seriously take into account those factors when they plan for retention strategies. Overlooking such factors can cause a loss to the service providers. The small difference that we see between the cumulative hazard rate of different levels of each of the covariates should not be underestimated, as often, the service providers have multimillion customers; 1% difference in the hazard rate could cause a loss of thousands of customers. We recommend further studies on the topic of customer survival time in the ISP industry by using parametric methods as the customer survival data in this industry is characterized by a high degree of censoring.

References

[1]   Blattberg, R. & Deighton, J., Manage marketing by the customer equity test, Harvard Business Review, 74(4), pp. 136-144, 1996.
[2]   Kim, W. G. et al., The effect of consumer-based brand equity on firms' financial performance, Journal of consumer marketing, vol. 20, 2003.
[3]   Rust, R. et al, Driving customer equity, The Free Press, 2000.
[4]   Bell D., et al, Seven barriers to customer equity management, Journal of service research, 2002.
[5]   Mohammed, Z. and Kotze, D., Survival data mining usefulness and complications, Data mining, text mining and their business applications and WIT Transaction on Information and Communication Technologies, Vol(35), WIT press, 2005.
[6]   Lu, J., Modelling customer lifetime value using survival analysis, SUGI (28), pp 120-128, 2003.
[7]   Cox, D. R. and Oakes, D., Analysis of survival data, Chapman & Hall, 1984.
[8]   Hougaard, P., Analysis of multivariate survival data, Springer, 2001.
[9]   Kleinbaum, D. G, Survival analysis, Springer, 1995.
[10]   Oakes, D., Biometrika Centenary: Survival analysis, Biometrika, 88(1), pp. 99-142, 2001.
[11]   Smith, P.J., Analysis of failure and survival data, Chapman &Hall/CRC, 2002.
[12]   Informa UK Ltd, Churn – Maximizing the profit through customer retention, 2001.

[13]   Nelson, W., Theory and application of hazard plotting for censored failure data, Technometrics, 14, 945-966, 1972.
[14]   Aalen, O. O., Nonparametric inference family counting processes, Annals of Statistics, 6, 701-726, 1978.
[15]   Kaplan, E. L., and Meier, P., Nonparametric estimation from incomplete observation, Journal of the American Statistical Association, vol. 80, 1958.