

PREDICTING MULTI-COHORTS' CUSTOMER RETENTION USING LIMITED INFORMATION

Apostolos Giovanis

Technological Educational Institute of Athens
Dept. of Business Administration
Agiou Spiridonos Street, Aigaleo, P.C.122 10, Athens, Greece
a.giovanis@wind.com.gr

Christos C. Frangos

Technological Educational Institute of Athens
Dept. of Business Administration
Agiou Spiridonos Street, Aigaleo, P.C.122 10, Athens, Greece
cfragos@teiath.gr

Dimitris Papathanasiou

SPSS BI Greece S.A.
Panormou & Karystou 5 Street, P.C. 115 23, Athens, Greece
dpapathanasiou@spss.gr

Abstract

Customer churn and retention rates are considered as the most important indicators to CRM activities such as modelling customer valuation and budget allocation. Such models, which are trying to predict customer's survival rate, assume that churn rates are constant both over time and across customers. The current practice, however, have shown that these assumptions are not valid and are currently tested in a systematic manner. The practice shows that churn propensity vary dramatically across customers and over time; yet, most models do not account for such differences. This study takes under consideration a number of factors that may underlie customer retention patterns. Specifically a model of mobile services' retention is developed that takes into account: duration dependence, cross-cohort effects, customer heterogeneity, and contractual obligation type. Moreover, a methodology is proposed that easily allows practitioners to evaluate the relative importance of each of these effects based on in-sample and out-of-sample performance using limited information. Using a dataset from a major Greek telecommunication company concerning customers' gross additions for twelve months and the relevant disconnections for the same time period ahead, two models on aggregate data at the monthly level are tested. The proposed modelling approach reveals that the notion of a constant churn rate is rejected. The representation of customer's retention process requires all the proposed components listed above towards improving the retention and churn rate forecasting performance.

Keywords: Retention, heterogeneity, cross-cohorts effects, latent class survival models, CRM

1. Introduction

A key characteristic of a contractual business setting is that customers' defection is observed. For example, the customer has to contact the firm to cancel his/her cell phone contract. In this business setting, practitioners are using several metrics such as retention rate for period t , which is defined as the proportion of customers active at the end of period $(t-1)$ who are still active at the end of period t , and churn rate for a given period which is defined as the proportion of customers active at the end of period $(t-1)$ who dropped out in period t .

In order to understand the nature of customer behavior in a contractual setting, it is useful to draw on survival analysis. One particularly useful concept for characterizing the distribution of customer lifetimes is that of the survival function, denoted by $S(t)$, which is the probability that a customer has "survived" up to time t (i.e., is still active at t). Recalling the definition of retention rate, it follows that,

$$S(t) = \prod_{i=1}^t r_i \quad (1)$$

which means that,

$$r_t = \frac{S(t)}{S(t-1)} \quad (2)$$

In a contractual setting, the empirical survivor function $S(t)$ is simply the proportion of customers acquired at time 0 who are still active at time t . A major problem in using the empirical survivor function to compute expected tenure or lifetime value is that the observed time horizon is often quite limited. Suppose a particular cohort of customers over their first twelve months with the firm are observed, which implies we $S(1), S(2), \dots, S(12)$ can be computed. The quantity $S(1) + S(2) + \dots + S(12)$ is the expected customer lifetime for the members of the cohort over this period. However, the expected tenure of a new customer would be underestimating, as the remaining life of those customers who are alive at the end of the twelfth month would be ignoring. In order to compute the true expected tenure, we need to be able to project the survivor function beyond the observed time horizon. That is, we need to create estimates of $S(13), S(14), \dots$ given the data $S(1), S(2), \dots, S(12)$. This projected survivor function is also needed if we wish to compute the expected residual tenure value of an individual who has been a customer for, say, two years.

The objective of this paper is to present an approach to the problem of projecting the survivor function for customers belonging to different cohorts, using a parametric duration model. The probabilistic model of contract duration is formulated based on customers' behavior incorporating at the same into the model some effects that influencing customer retention pattern. Such effects are customer heterogeneity accounting for unobserved differences across customers, cross-cohort effects accounting for differences that may exists between customers who subscribed in different periods (cohorts), and finally the type of contractual obligation effect accounting for the type of acquisition offers received by the customer in order to purchase the service. The resulting model offers useful diagnostic insights and is very easy to implement using Microsoft Excel.

In the next section, the probability model of customer lifetime is presented and the value of using a formal model to predict future customer behavior is demonstrated. Several issues that arise from the study are discussed at the end.

2. Model Development

To forecast the number of subscribers retaining service in future periods, a parametric form for the "survival" probability, $S(t)$ (i.e., the probability that a customer has maintained service until time t) is proposed. That is, after calibrating the model on n periods of data, the likelihood of maintaining service until time t can be forecasted by calculating $S(t)$ for any $t > n$. The availability of survival probability provides the opportunity of estimating other related measures of interest, such as, the predicted number of customers who still have the service at time t , $(N_0 \times S(t))$, where N_0 are the number of customers who acquired at time 0 and $(N_0 \cdot [S(t) - S(t-1)])$ the number of customers expected to drop service between time t and $t + 1$.

The conditional rate of defection given that the customer is still "alive", known as hazard rate, will be utilized as base for the model development. A latent class proportional hazard model, able to account for various effects, will be used in order to incorporate the previously described effects into the proposed model. We do so within a mixture model specification to incorporate heterogeneity, in which the probability that a randomly selected customer has not dropped service by time t is given by

$$S(t) = \int S[t|\theta_i, b, X(t)] \cdot g(\theta_i) d\theta_i \quad (3)$$

where θ_i is an individual-specific set of latent parameters, $X(t)$ is a vector of covariates at time t , and b is the effect of these covariates. Eq. (3) could be considered as a mixed-effects hazard model with both fixed and random components. The model in Eq. (3) consists of two main components: $S[t|\theta_i, b, X(t)]$ and $g(\theta_i)$.

The term $S[t|\theta_i, b, X(t)]$ specifies the probability that a consumer maintains the service until time t , which can be written in terms of its hazard function as:

$$S[t|\theta_i, b, X(t)] = \exp \left[- \sum_{v=1}^t \left(\int_{v-1}^v h(u|\theta_i, \beta, X(t)) du \right) \right] \quad (4)$$

Rather than assuming that all subscribers are homogeneous, the mixing distribution $g(\theta_i)$ allows for unobserved differences in subscribers' tendencies to discontinue service, because some may be inclined to do so after only a few months whereas others may be more reluctant.

Next, we describe how each of the four effects that are considered is formulated within the mixture model presented in Equations (3) and (4).

2.1. Duration Dependence

The likelihood that a subscriber will drop service may change according to the length of time for which he or she has had it. Therefore, the Weibull distribution for the baseline hazard is adopted, which is flexible and commonly used in proportional hazards models (Morrison & Schmittlein 1980; Seetharaman & Chintagunta, 2003):

$$h_0(t|\lambda, c) = c \cdot \lambda \cdot t^{c-1} \quad (5)$$

The Weibull distribution nests the “strawman” exponential distribution when $c = 1$; if subscribers do not exhibit duration dependence in their likelihood to drop service, the model collapses to a constant hazard rate λ . Values of $c > 1$ yield an increasing hazard rate, implying that subscribers are more likely to discard service the longer they have had it. Conversely, $c < 1$ leads to a decreasing hazard rate, such that subscribers become less likely to discard service as their tenure increases. Depending on the nature of duration dependence, managers may want to allocate their marketing efforts toward “older” or “younger” subscribers.

2.2. Cross-Cohort Effects

The baseline hazard function given in Eq. (5) is independent of the time at which subscribers began service (or the cohort to which they belong to). To allow for systematic differences across cohorts, we incorporate a cohort-specific effect (Fader and Hardie, 2008; Schweidel et al., 2008)

$$h_0(t|\lambda, c) = c \cdot \lambda \cdot \exp(\phi(j-1)^\tau) \cdot t^{c-1} \quad \text{for } j = 2, 3, 4, \dots \quad (6)$$

where t is the time that has elapsed since customers began service in month j of the observation period. The parameter ϕ governs the direction of the shift relative to the first cohort, where $\phi > 0$ implies that other cohorts have higher hazard rates relative to the first observed cohort, while $\phi < 0$ implies that other cohorts have lower hazard rates comparatively. If $\phi = 0$, then all cohorts have the same baseline hazard rate. The magnitude of the shift on cohort j is affected by the parameter τ , where $\tau > 0$ suggests that the magnitude of the shift increases with j (cohort), while $\tau < 0$ suggests that the magnitude decreases. For $j=1$, we use the baseline hazard rate given in (5), implying that the cross-cohort effect can be interpreted as a shift in the baseline hazard rate relative to the baseline hazard rate of the first cohort, where the ratio of the baseline hazard rates is:

$$q(j, \phi, \tau) = \frac{h(t, j|\lambda, \phi, \tau)}{h(t, 1|\lambda, \phi, \tau)} = \exp(\phi(j-1)^\tau) \quad \text{for } j = 2, 3, 4, \dots \quad (7)$$

This simple two-parameter model component, as given in (7), should be sufficient to capture (or at least adequately approximate) any cross-cohort dynamics present in our dataset (Schweidel et al., 2008).

Combining the cross-cohort effects with the individual-level baseline hazard function from Eq. (4) yields the conditional probability of a subscriber from cohort j of a particular service maintaining it until time t given by

$$S(t, j|\lambda, \Theta) = \exp\left(-\sum_{v=1}^t \int_{v=1}^v c \cdot \lambda \cdot q(j, \phi, \tau) \cdot u^{c-1} du\right) \quad (8)$$

$$S(t, j|\lambda, \Theta) = \exp(-\lambda \cdot t^c \cdot \exp(\ln(q(j, \phi, \tau)))) \quad (9)$$

2.3. Subscriber Heterogeneity

The likelihood of dropping a particular service is not the same for all customers. In order to model such a process a mixture distribution is specified, which accounts for unobserved subscriber heterogeneity in the likelihood of dropping a particular service. If all subscribers have the same propensities for discarding a service, the mixing distribution would simply be a spike [$g(\theta_i) = \lambda$], but this would be an extremely restrictive assumption. Instead, we allow for heterogeneity across service subscribers by assuming that each subscriber's λ_i is drawn from a gamma distribution:

$$g(\lambda|r, a) = \frac{a^r \cdot \lambda^{r-1} \cdot e^{-a\lambda}}{\Gamma(r)} \quad (10)$$

The gamma distribution is chosen not only for its flexibility, but also because it is the conjugate prior for the Weibull distribution and therefore is commonly used as a mixing distribution for this purpose (e.g., Follman and Goldberg 1988). As such, the marginal probability of a customer in cohort j surviving until time t is found by integrating (9) over the mixing distribution (10), as shown in detail in (3), and is given by,

$$S(t, j | r, a, c, \phi, \tau) = \left(\frac{a}{a + t^c \cdot \exp(\phi \cdot (j-1)^\tau)} \right)^r \quad (11)$$

2.4. Contractual Obligation Type

We postulate the existence of discrete segments of customers, where within each segment, individual level time-to-churn is Weibull-distributed with parameters λ_s, c_s . λ_s is distributed across segment members according to the Gamma distribution with parameters r_s and a_s . In the specific study we assume the existence of two segments expressing customers who received terminal subsidization for service activation (e.g. 12 month contract obligation) vs. customers who preferred to take the subsidy as bill discount (they are allowed to switch any time they wish). The parameters expressing the cross-cohort effect are assumed to be the same for both segments since for all cohorts the subsidization policy was the same. Based on the above the probability that a randomly selected customer has not dropped service by time t is given as a finite mixture of survival functions given by:

$$S(t, j | r, a, \phi, \tau) = 1 - \sum_{i=1}^2 p_i \cdot (1 - S(t, j | r_i, a_i, \phi, \tau_i))$$

$$\sum_{i=1}^2 p_i = 1 \quad (12)$$

3. Empirical Analysis

The data used in our empirical analysis were provided by a major telecommunications operator that offers a broad range of mobile services to its customers. To evaluate the importance of the four factors, we fit two models to the data the company offered. The first model (named WGCC) does not consider the contractual obligation type determined by the acquisition offers, while the second model named (WGCC-2S) considers all four effects described above. The models calibration and forecasting performance are tested using the Bayesian Information Criterion (BIC)

$$BIC = -2 \cdot LL + q \cdot \ln(D) \quad (13)$$

and the Mean Absolute Percentage Error

$$MAPE = \frac{1}{k} \sum_{i=t+1}^{t+k} \left| \frac{N_t - E_t}{A_t} \right| \quad (14)$$

respectively, where q is the number of parameters, D the number of observations, N_t the actual number of customers survived up to time t , E_t the expected number of customers survived up to time t and k the forecasting horizon.

The company provided subscription information from Sept 2006 to September 2008, indicating (in aggregated monthly data) the number of customers subscribing at the end of each month. Thus, the first group that we observe from the time that it begins service is the Oct 2006 cohort. The company noticed that around 50% of the subscribers were received terminal subsidization in order to activate the service. The chosen models calibration period is 12 months e.g. up to Sep 2007. Then using the parameters provided during the model calibration the models' forecasting ability is tested in terms of service retention numbers from the period between Oct 2007 and Sep 2008. Figure 1 illustrates the empirical survival function for the available data considering the full time period of 24 months, while Figure 2 provides the relevant empirical hazard function for the same time period.

From the customers that maintain service and those that churn during the calibration period, we can construct the log-likelihood of the observed behavior of customers from cohort j :

$$LL(\Theta|j) = N_{T-j+1,j} \cdot \log(S(T-j+1, j)) + \sum_{t=1}^{T-j} [(N_{t,j} - N_{t+1,j}) \cdot \log(S(t, j) - S(t+1, j))] \quad (15)$$

where the first term accounts for customers that maintain service through the entire calibration period and the second term accounts for households that churn during the calibration period, summing over all intervals in which the customers could discard service. The log-likelihood for the full data set, consisting of all cohorts that began service during the calibration period, is as follows:

$$LL(\Theta) = \sum_{j=1}^{T-1} LL(\Theta|j) \quad (16)$$

4. Results

For models' calibration, the Solver Add-In of MS Excel was used. A procedure provide by Billo (2001) was utilized to obtain the standard deviations of the coefficients. Table 1 presents the results of models' calibration and forecasting performance. Both models parameters are statistical significant. With respect to model WGCC formulation, it can be seen that the value of parameter c is greater than 1, implying that subscribers are more likely to discard service the longer they have had it. The values of parameters ϕ and τ imply decreased hazard for all cohorts relative to $j=1$ with an increasing effect with respect to j . The second model, which considers acquisition type offers in order to construct a two latent class model, presents a better calibration and forecasting performance (lower BIC and MAPE indices). WGCC-2S formulation implies that the provided customer base consists of two segments which have different behavior in terms of churn propensity. The first segment accounting for 50,6% of the population are more likely to discard service the longer they have had it, while the other 49,4 % are more likely to retain service as the time goes by. In terms of cross-cohorts effects the values of model's parameters imply that all cohorts presents decreased hazard relative to $j=1$ but the effect diminishes with j . The fitting and forecasting performance of the two models are illustrated in Figures 1 and 2 which presents the aggregate survivors and churners over time respectively. It's obvious that WGCC-2S provides a better representation of customers' decision process in terns of retention/churn propensity.

5. Conclusions

The problem of predicting customer retention rate is very vital for organizations. A reliable forecasting methodology of this customers' measure provides better estimations of customer lifetime value (CLV) which in turns is used for customer segmentation and CRM budget allocation. The current study proposed a two latent class finite mixture probability model aiming to represent customers' decision mechanism in terms of retain or discard a mobile service over time. The proposed model includes four external effects which affect the customers' behavior represented by quantities such as P(alive), mean residual lifetime and (residual) CLV, all conditional on the customer's observed behavior, which can be used for policy development. The proposed methodology could be implemented very easily in spreadsheet environment and become a very good tool for CRM policy development.

References

1. Billo, J. (2001). Excel for Chemists: A Comprehensive Guide. John Wiley & Sons Inc.
2. Fader, P. and Hardie, B. (2009). Probability Models for Customer-Base Analysis. Journal of Interactive Marketing 23: 61–69
3. Follman, D.A. and Matthew S.G. (1988). Distinguishing Heterogeneity from Decreasing Hazard Rates. Technometrics 30 (4): 389–96.
4. Morrison, D.G. and Schmittlein, D.C. (1980). Jobs, Strikes, and Wars: Probability Models for Duration. Organizational Behavior and Human Performance 25(2): 224–51.
5. Schweidel, D. Fader, P. & Bradlow, E. (2008). Understanding Service Retention Within and Across Cohorts Using Limited Information. Journal of Marketing 72: 82–94
6. Seetharaman, P.B. and Chintagunta, P.K. (2003). The Proportional Hazard Model for Purchase Timing: A Comparison of Alternative Specifications. Journal of Business and Economics Statistics 21 (3): 368–382.

Appendix

Tables

Table 1: Models Parameters and Performance Measures for the two Modes

Parameters	WGCC		WGCC-2S	
	Estimate	St. Error	Estimate	St. Error
a_1	0,0631	0,0018	1,6890	0,0325
r_1	0,0662	0,0034	0,2816	0,0122
c_1	6,2031	0,7809	4,3154	0,5340
ϕ	-0,6977	0,0299	-0,3583	0,0104
$T\alpha$	0,0729	0,0192	-0,0319	0,0109
a_2			4,5373	0,2259
r_2			1,2487	0,1213
c_2			0,4067	0,0141
P			0,5061	0,0085
LL	-255.800,06		255.214.64	
BIC	511.622,63		510.469,77	
$MAPE$	0,744%		0,038%	

Figures

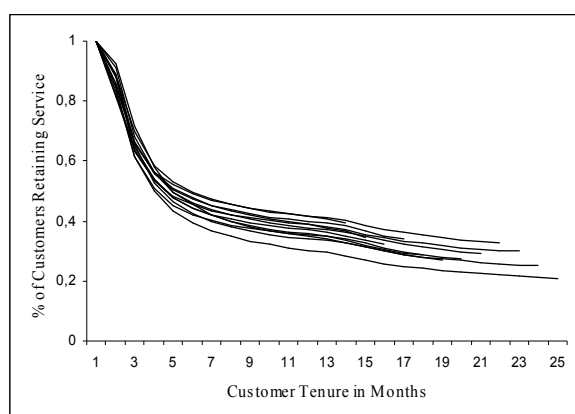


Figure 1: Cohort-Level Retention Behavior

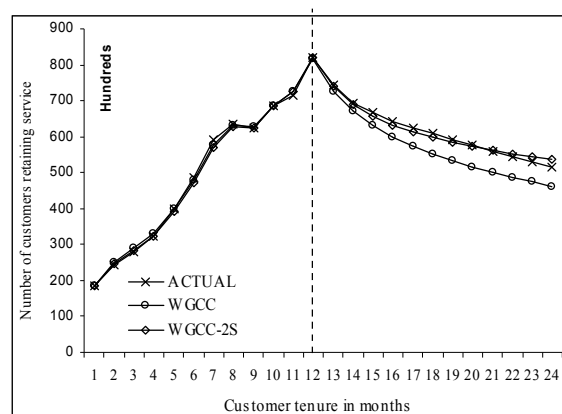


Figure 3: Actual vs. Projected Aggregate Number of Survivors

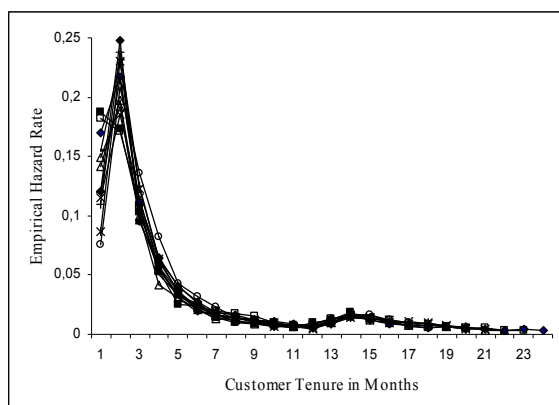


Figure 2: Cohort-Level Empirical Hazard Rate

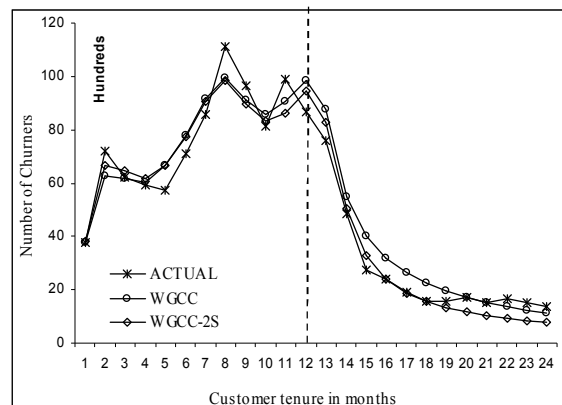


Figure 4: Actual vs. Projected Aggregate Number of Churners