Les Cahiers du GERAD ISSN: 0711–2440

Will this Gold customer remain Golden this year?: Intra-Periodic Forecasting of End-of-Period Customer Status Using Nonhomogeneous Poisson Processes with Random Effects

M. Fredette, J.-F. Ouellet

G–2008–52

July 2008

La publication de ces rapports de recherche bénéficie d'une subvention du Fonds québécois de la recherche sur la nature et les technologies.

Will this Gold customer remain Golden this year?: Intra-Periodic Forecasting of End-of-Period Customer Status Using Nonhomogeneous Poisson Processes with Random Effects

Marc Fredette

GERAD and
Service de l'enseignement des méthodes quantitatives de gestion
HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada, H3T 2A7
marc.fredette@hec.ca

Jean-François Ouellet

Service de l'enseignement du marketing
HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada, H3T 2A7
jean-francois.ouellet@hec.ca

July 2008

Les Cahiers du GERAD

G–2008–52

Abstract

Top-tier customers—that is, those 20% of customers that typically bring in 80% of all profits—are extremely valuable to companies. In the many instances in which organizations attribute top-tier status to customers based on their consumption behaviour within a specific period, such as a year, it becomes very important to determine, during this period, how likely those gold customers are to retain their top-tier status going into the next period. This allows better planning at the corporate level, but can also allow for corrective measures or special retention efforts to be deployed. However, while models exist to predict customer churn or customer lifetime value either at the beginning of a period or on a continuous basis according to the evolution of inter-purchase time, no model allows for a continuous re-estimation of customer status or value according to calendar time, based on historical data and year-to-date information. To address this problem, we develop a model of intra-periodic forecasting of customer behaviour that uses nonhomogeneous Poisson processes with random effects. We then empirically assess the performance of this model using data from the loyalty program at a major commercial airline.

Key Words: Repeat buying; probability models; forecasting; loyalty marketing.

Résumé

Les clients de haut-niveau, c'est-à-dire les 20% de clients qui amènent généralement 80% des profits, sont extrêmement utiles aux entreprises. Dans les nombreux cas où les organisations attribuent des statuts de haut-niveau à leurs clients sur la base de leurs habitudes de consommation annuelle, il devient très important de déterminer, au cours de l'année, la probabilité qu'un client conserve un tel niveau pour l'année suivante. Cela permet non seulement une meilleure planification au niveau de l'entreprise, mais permet aussi d'identifier des mesures de maintien plus appropriées.

Dans cet article, nous proposons ensuite un modèle de prévision du comportement de la clientèle à l'aide de processus de Poisson non-homogènes avec effets aléatoires. Ensuite, nous évaluons empiriquement l'adéquation de ce modèle en utilisant les données d'un programme de fidélité d'une grande compagnie d'aviation commerciale.

1 Introduction

It is rather well accepted in the academic and business literature that around 20% of any firm's customers bring in some 80% of corporate revenues. Past research has indeed confirmed this Pareto-like distribution of customer lifetime value across various segments. For instance, in their research conducted in the airline industry, Rust, Lemon, and Zeithaml (2004) show that only 11.6% of customers at a leading U.S. airline have a lifetime value of $500 or more and that this segment makes up approximately 50% of that airline's total customer equity. Popular marketing gurus contend that these highly lucrative customers should not be targeted with marketing activities as they are typically "maxed-out," using every service they could ever need, and potentially already giving their provider a near-100% share-of-wallet. These customers should rather be given as high a level of service as possible, while marketing efforts should rather be targeted at the middle 60% of customers which bring in a share of profit that could be increased (e.g., Reichheld and Teal 1996; Reichheld 1993; Reichheld and Sasser 1990).

This being said, any segment cohort, when followed through multiple periods, will show decline as customers move up or down lifetime value segments, or altogether defect (Drozdenko and Drake 2002). Considering the importance of so-called "Gold Customers" for every organization, predicting the activity levels of these customers, and figuring out whether they will remain top-tier customers over the next period should hold considerable appeal. Indeed, such prediction would allow better planning and allocation of resources across segments and customer tiers, and drive the overall strategy as to whether it should focus on retaining top-tier customers or acquiring new ones. In this context, several general approaches have been developed to model customer behaviour, most of which stemming from research pertaining to the evaluation and forecasting of customer lifetime value (e.g., Fader, Hardie, and Lee 2005a; 2005b; Borle, Singh, and Jain 2008; Venkatesan, Kumar, and Bohling 2007; Simester, Sun, and Tsitsiklis 2006; Rust, Lemon, and Zeithmal 2004). Under this paradigm, the focus lies on evaluating the total number of transactions or revenues per customer by the end of a specific period in order to steer investments towards certain specific customers as a way to manage customer equity. Although not specifically developed with the top-tier customer in mind, these models could still be used to determine, at the beginning of a new period, whether each customer is most likely to remain a "Gold customer," or to move down or even defect.

While these models perform generally well and could provide significant benefits in the evaluation and forecasting of top-tier customer "downward-migration," their contribution to other down-to-earth and pragmatic managerial concerns and marketing preoccupations is rather limited. For example, an advertising manager might be interested in knowing which top-tier customers are most at risk of loosing top-tier status, allowing better allocation of resources to better reach these customers. Given that access and responsiveness are key determinants of perceived service quality (Parasuraman, Zeithaml, and Berry 1985; 1991), a customer service director may also want to know when a surge in telephone orders from top-tier customers is likely to occur in order to better plan for the availability of customer service personnel. A product manager might be interested in finding out at which period demand is likely to be higher or lower in order to manage pricing differentially across sub-periods to capture as much of the value as possible. A customer relationship manager in a loyalty program at an airline may want to know in October what is the likelihood of a Gold Customer to have accrued enough air miles to remain a Gold Customer going into the next year in order to help plan end-of-year special offers. Or, a key account executive may want to know whether a client is to remain a "Gold Customer" this year despite a low number of transactions over the first three quarters, and whether he or she should trigger special recovery efforts. In this last case, considering that account managers' compensation is—in some 92% of surveyed U.S. corporations—based on reaching annual sales quotas (Ryals and Rogers 2005), being able to assess, say, in August whether target sales for a given key account are likely to be met by the end of December should hold considerable importance. This is especially true in making decisions regarding retention-oriented activities—that is, service—or acquisition-oriented activities—e.g., cross-selling, up-selling, solicitation, etc.—even in those top-tier segments the academic literature argues we should really only try to serve well and not waste our marketing resources targeting.

Under the hood, existing models consider the flow of transactions to be homogeneous within the period. For instance, a forecast of 156 transactions per year for a given customer is implicitly considered to be 13 transactions a month, or 3 transactions a week, and thus does not take into account the fact that these transactions can be distributed otherwise within the year. However, in many product categories, sales are affected by cyclical events like seasons or holiday periods. Alternatively, these existing models could be used to forecast the level of transaction in a specific period—for instance, the month of March, or week 28—for each customer. In this case, data about a specific customer's actions in the months of March of the 2, 3, 4, or more years before would be used to predict this customer's actions over the next month of March. Because it relies on past data for comparable periods, this approach yields two major drawbacks: (1) Due to the usually limited amount of historic data (e.g., companies preserve customer data for a limited period of time, oftentimes for 3 years or less), it greatly extends prediction intervals, oftentimes past a reasonable and managerially useful threshold; and (2) it considers these periods to be isolated from one another—that is, it fails to take into account the effects of preceding (or future) periods. For example, the recency of a given transaction can affect repurchase in the next period either positively (e.g., Fader, Hardie, and Lee 2005a) or negatively (e.g., it is less likely that someone who booked a holiday cruise this week will book one again next week (Venkatesan, Kumar, and Bohling 2007)).

To overcome these problems and address these managerial preoccupations, we develop a model allowing finite-horizon prediction of recurrent events using flexible nonhomogeneous Poisson processes with random effects. We first review the various models that were introduced so far in the marketing literature and comment on their drawbacks with regards to addressing the problems stated above (§2). In §3, we outline the assumptions relative to the use of mixed nonhomogeneous Poisson models for predicting recurrent events before providing details on the development and estimation of our model, in §4. This is followed by an empirical analysis of model performance using data from a major airline company (§5). We finally conclude with a discussion of our results, limitations, and avenues for future research.

2 Forecasting Customer Activity

The extant literature provides several models to predict individual and aggregate transactions or unit sales within a certain future period, as well as the timing of these transactions. Early contemporary models focused on the evaluation of the potential value of a customer as the product of the probability that a customer purchases times the profit margin of that purchase

(Verhoef and Donkers 2001; Lewis 2005). The limited applicability of this approach in non-contractual settings, where the expected purchase pattern of the customer is more or less stable, soon prompted the development of alternative models.

More recent models allowed for a more or less precise prediction of expected consumption (e.g., number of transactions to be made) according to past purchases. Rust et al. (2004) computed customer lifetime value as a function of frequency of category purchases, average quantity of purchase, and brand switching patterns combined with the firm's contribution margin. However, an important limitation of this approach was that the resulting model was inherently static. Moreover, this research, along with similar work by Gupta and Lehmann (2003), assessed the average value of a customer at some level of aggregation (or customer segmentation). That is, these models did not provide individual customer-level insights, a key objective in our research. Although alternative models take into account heterogeneity between various customers (e.g., Fader, Hardie, and Lee 2005b), and also offer better prediction for defection, or zero-purchase behaviour (e.g., Batislam, Denizel, and Filiztekin 2007), they usually predict either time-to-next-purchase and/or time between purchases to derive a global assessment of behaviour for each customer (Borle, Singh, and Jain 2008; Venkatesan, Kumar, and Bohling 2007), which brings in further problems.

In relation with our objectives, the problem with these models, and especially with the NBD–Pareto model and its variants (Fader, Hardie, and Lee 2005b), is that it assumes that customers buy at a steady rate for a period of time and then become inactive—that is, they do not take into account seasonality, cyclicality, or otherwise increases or decreases in consumption behaviour and thus focus on customer status as either active or inactive. This is hardly the case in many product categories, and as Figure 1 shows, certainly not the case with flying patterns of top-tier customers in the airline industry. In addition, these models tend to focus on predicting end-of-period outcome without explicitly taking into account behaviour-to-date in the period of interest. In fact, only Borle, Singh, and Jain's (2008) model observes complete customer lifetimes, updating expected customer lifetime value and probability of defecting (i.e., the hazard rate) after each purchase, given that the customer has survived until a particular purchase occasion. However, in this latter case, only inter-purchase time updates customer lifetime value predictions, not calendar time.

Other deterministic models have also been introduced to predict various customer states, such as active/inactive (in the case of churn models), and typically rely on logistic regression, hazard models, or Markov chains (Bhattacharya 1998; Bolton 1998; Pfeifer and Carraway 2000; Van den Poel and Larivière 2004; Buckinx and Van den Poel 2005; Simester, Sun, and Tsitsiklis 2006; Schweidel, Fader, and Bradlow 2008). Here again, however, these approaches are all inherently static and do not shed light on intra-periodic questioning with regards to the direction a specific top-tier customer is actually headed toward.

3 Prediction of Recurrent Events with Mixed Poisson Models

3.1 Mixed Nonhomogeneous Poisson Processes

The objective of our work is therefore to accurately predict whether a currently top-tier customer will have cumulated enough transactions by the end of the current period to qualify again for top-tier status over the next period. That is, there is a finite population of unit customers with top-tier status at the beginning of the year ($i = 1, \ldots, k$), and we wish to predict the number of transactions to be made by these individual consumers and if so desired, from there, the aggregate number of such transactions across the population, or across a given segment. Specifically, we consider what we term finite-horizon prediction, the objective of which is to predict the total number of events (i.e., transactions), for an individual or the whole population over a specified time period $(0, T)$ on the basis of events that have already occurred up to given times $t_i \leq T$ for the units in the population. In practice, this interval $(0, T)$ would typically refer to a calendar or fiscal year time period, and the various t_i's would usually take the same value for all units given that calendar year begins at the same time for everybody.

The context of this research is frequent flyer status within a specific airline loyalty program, where top-tier, "Gold" status is obtained after having flown 20 different flights in a calendar year. Let t represent the number of days elapsed since the beginning of the calendar year, and let $N_i(u, v)$ denote the number of flights taken in the age interval $u < t \leq v$. The objective is then to predict $N_i(0, T)$ where $T = 365$ on non leap years—that is, the total number of flights flown by customer i between the beginning and the end of the calendar year. Of course, as an added benefit, it may eventually be useful to predict all flights flown by all customers during the whole year by predicting

$$N_+ (0, T) = \sum_{i=1}^{k} N_i (0, T) . \tag{1}$$

Of course, $N_i(0, T)$ will ultimately be known for each i once the calendar year is over. However, in several situations such as the various instances described in introduction, it might be useful to predict this value on the basis of previous experience with this customer but also on the basis of flights already taken during that calendar year. For convenience, we consider continuous time processes where two events cannot occur simultaneously. From this point on, we also write $N(t)$ for $N(0, t)$. Different types of recurrent events processes are discussed in the literature on point processes (Grandell 1997). These are all characterized by an event intensity function

$$\lambda(t| H(t)) = \lim_{\Delta t \to 0} \frac{P[N(t, t + \Delta t) = 1| H(t)]}{\Delta t} \tag{2}$$

where $H(t)$ denotes the history of the process up to time t. Poisson processes are Markovian because (2) depends only on t. The intensity, or rate, function is then simply denoted by $\lambda(t)$, and

$$N(t) \sim PP(\lambda(t))$$

means that $N(t)$ is a nonhomogeneous Poisson process (NHPP) with rate function $\lambda(t)$.

Figure 1 provides a plot of the number of flights flown each day for the first 3 years of data that we have, for those with top-tier frequent flyer membership after the first year. This graph clearly shows the seasonal (that is, nonhomogeneous) character of the flying habits of top-tier frequent flyers. It also shows, as is expected from the marketing literature pertaining to loyalty programs, a number of flights flown daily that diminishes with time as members from this top-tier cohort leave the program and/or the company, or change flying habits.

It is well known that in a Poisson process, the total number of events over any interval has a Poisson distribution, and that the number of events $N(s_1, t_1)$ and $N(s_2, t_2)$ in two

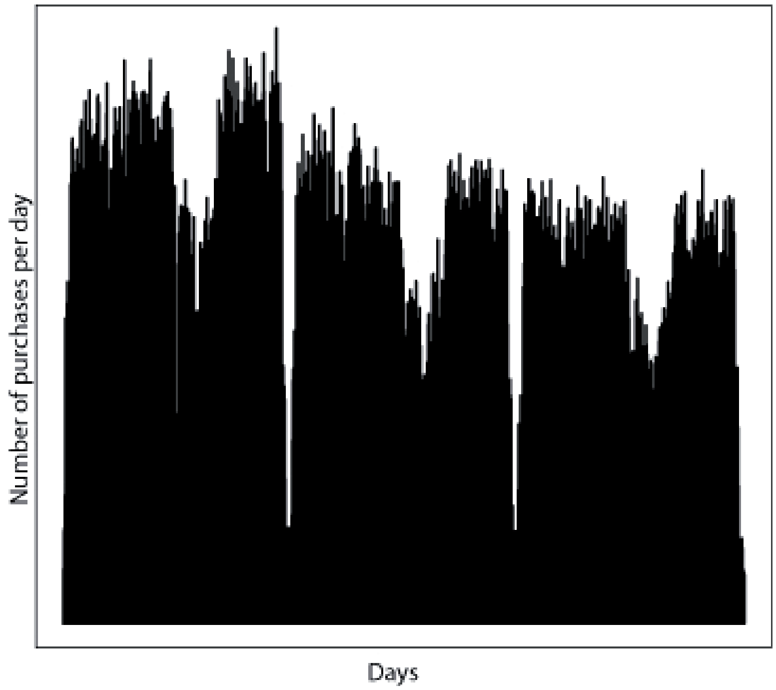

Figure 1: Total number of purchases per day over 3 years

nonoverlapping time intervals (s_1, t_1) and (s_2, t_2) are independent. These two properties make Poisson processes easy to use with prediction problems involving recurrent events. These have been applied to a wide range of business settings, both in marketing (e.g., Wang et al. 2007) and non-marketing contexts (e.g., Teunter and Klein Haneveld 2008). However, in populations with heterogeneous units, it is generally necessary to extend the models by including unit-specific random effects. Such models are termed random-effects, or mixed, Poisson processes (e.g., Lawless 1987; Grandell 1997).

We model the rate function for a single process with parametric forms $\lambda(t; \alpha, \beta) = \alpha f(t; \beta)$, where α is a scalar and β is a vector of low dimension. This parameterization is convenient because $f(t; \beta)$ and α measure different aspects of a NHPP; the function $f(t; \beta)$ describes the shape of the rate function, and α represents the overall event frequency. In the finite-horizon problems, it is convenient to choose α so that $\mathrm{E}[N(0, T)] = \alpha$, in which case $\int_0^T f(t; \beta) dt = 1$. That is, $f(t; \beta)$ has the form of a probability density function over $(0, T)$.

To consider scenarios in which heterogeneity is observed among the processes for different units, we incorporate unobservable *iid* random effects in our model. The model considered in this article is

$$\begin{aligned} N_i(t) | \alpha_i &\sim PP(\alpha_i f(t; \beta)), \qquad (3) \\ \alpha_i &\sim gamma(a, b), \end{aligned}$$

where $i = 1, \ldots, k$. The parameterization for the gamma distribution is such that $\mathrm{E}[\alpha_i] = a/b$ and $\mathrm{Var}[\alpha_i] = a/b^2$. A suitable function $f(t; \beta)$ for the problem at hand will be proposed in next section. Fredette and Lawless (2007) proposed a similar prediction model, but with a different function $f(t;\beta)$, to forecast automobile warranty claims.

3.2 Prediction

We seek to construct prediction intervals for a future random variable Y, given observed data $X = x$. Such intervals are of the form $(L(x), U(x))$, and we attempt to find intervals where $P[L(X) \leq Y \leq U(X)]$ equals some specified fixed value $1-\gamma$, in which case $(L(x), U(x))$ is called a $1-\gamma$ prediction interval (e.g., Barndorff-Nielsen and Cox 1996) and $1-\gamma$ is called its coverage probability.

In the context discussed in this article, we wish to use the information regarding the k processes that is available at a certain given time (that is, the number of customers who had already flown by a given day of the year) to make predictive statements about the remaining number of events to be observed. As it is the focus of our article, only the prediction of a single count $N_i(t_i, T)$ is discussed here, but extensions to predict the sum of all counts (1) are discussed in Appendix A.

For each process, the information available to make our prediction consists of the total number of events, $N_i(t_i)$, and the set of occurrence times, $\tau_i(t_i) = \{\tau_{i1}, \ldots, \tau_{iN_i(t_i)}\}$. Conditional on this information, each $N_i(t_i, T)$ has a negative binomial distribution with parameters $a + N_i(t_i)$ and $(b + F(t_i;\beta))/(b + F(T;\beta))$, where $F(t;\beta) = \int_0^t f(u;\beta)du$ (a proof is given in Fredette and Lawless 2007). This predictive distribution is hereinafter denoted as $NB(a + N_i(t_i), (b + F(t_i;\beta))/(b + F(T;\beta)))$, with probability function given by

$$P[N_i(t_i, T) = n | N_i(t_i); a, b, \beta]$$

$$= \frac{\Gamma(a + N_i(t_i) + n)}{\Gamma(a + N_i(t_i))n!} \left(\frac{F(T;\beta) - F(t_i;\beta)}{b + F(T;\beta)}\right)^n \left(\frac{b + F(t_i;\beta)}{b + F(T;\beta)}\right)^{a+N_i(t_i)}, \quad (4)$$

and an expectation given by

$$= (a + N_i(t_i)) \left(\frac{F(T;\beta) - F(t_i;\beta)}{b + F(t_i;\beta)}\right).$$

Note that the occurrence times do not appear in this distribution; only knowledge of $N_i(t_i)$ is needed to determine this conditional distribution. However, the occurrence times will enter into the estimation of model parameters β.

3.3 Plug-in Prediction Intervals

Let $N(t) = (N_1(t_1), \ldots, N_k(t_1))$ and $\tau(t) = \{\tau_{ij}; i = 1, \ldots, k; j = 1, \ldots, N_i(t_i)\}$. A prediction interval for $N_i(t_i, T)$ is an interval $[L(N(t), \tau(t)), U(N(t), \tau(t))]$ such that

$$P[L(N(t), \tau(t)) \leq N_i(t_i, T) \leq U(N(t), \tau(t)); a, b, \beta] = 1 - \gamma.$$

Such an interval is called an exact $1-\gamma$ prediction interval for $N_i(t_i, T)$. In most settings (including the one considered in this paper), one cannot find exact prediction intervals when the parameters a, b, and β are unknown. This is analogous to the non-existence of exact confidence intervals for parameters in most statistical models. The alternative is to find an interval with an approximate coverage probability of $1-\gamma$. This can be accomplished in one way by finding an interval $[L, U]$ such that

$$P[L \leq N_i(t_i, T) \leq U, \hat{a}, \hat{b}, \hat{\beta}] = 1 - \gamma, \tag{5}$$

where only $N_i(t_i, T)$ is treated as a random variable, and where $\hat{a}$, $\hat{b}$, and $\hat{\beta}$ are the maximum likelihood estimates (MLE's) obtained from the likelihood function based on the observed data, which is (Lawless 1987):

$$L(a, b, \beta | N(t), \tau(t)) = \prod_{i=1}^{k} \left[\left(\prod_{j=1}^{N_i(t_i)} f(\tau_{ij}; \beta \right) \left(\frac{b^a}{(b + F(t_i; \beta))^{a+N_i(t_i)}} \right) \left(\frac{\Gamma(a + N_i(t_i))}{\Gamma(a)} \right) \right]$$

The interval (5) is called a "plug-in" $1-\gamma$ prediction interval. Essentially, this method assumes that the true parameter values are in fact $(\hat{a}, \hat{b}, \hat{\beta})$ and thus ignores completely the uncertainty in $(\hat{a}, \hat{b}, \hat{\beta})$ relative to (a, b, β). When the observed data set is very large, so that $(\hat{a}, \hat{b}, \hat{\beta})$ can be assumed close to (a, b, β), then the coverage probability of this interval will be close to 1-γ. [1]

Plug-in prediction intervals with an approximate coverage probability of $1-\gamma$ can easily be obtained by calculating the $\gamma/2$ and the $1-\gamma/2$ quantiles based on the predictive probability function $P[N_i(t_i, T) = n | N_i(t_i); \hat{a}, \hat{b}, \hat{\beta}]$ given by (4). This "plug-in" predictive distribution function can also be used to make other interesting predictive statements:

- At time t_i, the probability that a customer will retain his "top-tier" status at the end of the year (i.e., in the context of our research, that this customer will have flown 20 flights) is estimated by:

$$\sum_{n=20-N_i(t_i)}^{\infty} P[N_i(t_i, T) = n | N_i(t_i); \hat{a}, \hat{b}, \hat{\beta}].$$

- At time t_i, a point prediction for the remaining number of events (i.e., flights to be taken) is given by:

$$\begin{aligned} &= (\hat{a} + N_i(t_i)) \left(\frac{F(T; \hat{\beta}) - F(t_i; \hat{\beta})}{\hat{b} + F(t_i; \hat{\beta})} \right) \\ &= \left(F(T; \hat{\beta}) - F(t_i; \hat{\beta}) \right) \left[\frac{\hat{a}}{\hat{b}} (1 - w_i(t_i)) + \frac{N_i(t_i)}{F(t_i; \hat{\beta})} w_i(t_i) \right], \end{aligned}$$

  where $w_i(t_i) = \frac{F(t_i; \hat{\beta})}{(\hat{b} + F(t_i; \hat{\beta}))}$.

We can see that four different factors derived from our prediction model (3), based on nonhomogeneous Poisson processes with random effects, are influencing the predictions being made:

[1] For smaller datsets, this method can be improved upon by calibrating the plug-in intervals obtained. See Lawless and Fredette (2005) for complete details.

1. The factor $F(T;\hat{\beta}) - F(t_i;\hat{\beta}) = \int_t^T f(u;\hat{\beta})du$ takes into account the estimated time nonhomogeneity over the interval $[t_i, T]$.
2. The ratio $\frac{\hat{a}}{\hat{b}}$ is an estimation, based on the whole sample, of the event frequency.
3. The ratio $\frac{N_i(t_i)}{F(t_i;\hat{\beta})}$ is an estimation, based on customer i, of the event frequency.
4. An increasing function $w_i(t_i)$. As t_i increases, the weight of the whole sample on the estimation of event frequency decreases while the weight of customer i's actual behaviour so far increases.

4 Predicting the number of flights taken by frequent flyers

Frequent flyers programs involve the systematic collection of detailed information regarding members' flying activities, thus allowing prediction of individual activity level based on the data already observed. The database at hand was obtained from the loyalty program of a major American commercial airline. It includes information on individual top-tier frequent flyers for a period of 3 years starting January 1^{st}, 2004 and provides, for each frequent flyer, a unique identifier along with the various dates that flights have been flown. To qualify for top-tier "Gold" membership, each frequent flyer had to fly at least 20 times over first calendar year—that is, between January 1^{st}, 2004 and December 31^{st}, 2004 inclusively.

The quantities we wish to predict are $N_i(366, 731)$ for each frequent flyer i—that is, the number of flights taken by each frequent flyer between the first and last day of the second calendar year of data. The dataset contains such data for 5,000 frequent flyers. In the second year, each of them had actually flown between 0 and 158 flights. Table 1 gives the distribution of total number of flights flown in year 2 for those who had qualified for top-tier membership at the end of year 1.

Table 1: Distribution of the number of flights taken over year 2 by frequent flyers who had qualified for top-tier status by the end of year 1

Number of flights flown	Year 2
0	94
1∼10	721
11∼19	1112
20+	3073

Early in a given year, the managers estimate, for the new year in progress, the eventual number of flights to be flown by each frequent flyers according to past years' data. Once data begin to accrue for each frequent flyer however, the methods in Section 3 can be used to predict the number of flights to be flown by each member, or the total flights to be flown for a group of, or all, frequent flyers.

We now propose to use model (3) to predict the total number of flights flown by a given top-tier frequent flyer over a calendar year. The choice of a suitable parametric form for $f(t;\beta)$ in (3) is crucial, because our predictions necessarily involve extrapolation into the future. Ideally, the shape of this function would be the same every year to reflect the periodicity of flying habits of frequent flyers. In addition, we would like to allow for a potential reduction of the amplitude of this function to reflect the fact that the number of flights flown usually diminishes over time.

We thus consider the function

$$f(t;\beta) = p(t-366;\beta_1,\beta_2) \times \exp\{C(d_t;\beta_3,\ldots,\beta_{K+3})\},$$

where:

- d_t is the day number of the year. For example, $d_1 = d_{366+1} = d_{366+365+1} = 1$ (the first year was a leap year). This will allow the function to retain the same shape year after year.
- At the beginning of the second year, we incorporate a decreasing proportion $p(.;\beta_1,\beta_2)$ to reflect the fact that some customers are likely to leave the program over time. Because of the obvious relationship between this phenomenon and a survival problem, we opted for a survival function $p(t-366,\beta_1,\beta_2) = S(t;\beta_1,\beta_2)$ such that $S(0;\beta_1,\beta_2) = 1$ and decreases thereafter. We used the Weibull survival function $S(t;\beta_1,\beta_2) = \exp\{(-t\beta_1)^{\beta_2}\}$ which is probably, along with the log-normal survival function, the most popular distribution for survival problems. We shall note that the value of this function has a useful interpretation. For example, when we fitted our prediction model using the information available at the middle of the 2$^{\text{nd}}$ year, we obtained that the value of the estimated survival function at the end of that second year was 80%. It means that we predict that the frequency of flights flown at the end of the 2$^{\text{nd}}$ year will be 20% smaller than what it was at the end of the 1$^{\text{st}}$ year.
- $C(t;\beta)$ is a cubic spline. Cubic splines are continuous piecewise cubic polynomials used in curve fitting. They have been found to have nice properties with good ability to fit sharply curving shapes (Harrel 2001). In order to use a cubic spline, we first have to determine an appropriate number of knots. Between each of these knots, the continuous function $C(t;\beta)$ is a cubic polynomial. Based on the data available after the first year, we found out that it was sufficient here to use $K = 4$ knots. In order to have approximately the same number of recurrent events between each knots, the knots are the 20%, 40%, 60%, and 80% quantiles of all the occurrence times observed that 1$^{\text{st}}$ year (i.e., 70, 140, 220, and 300 days). The explicit form of this piecewise cubic polynomial is given by:

$$\begin{aligned} C(t;\beta_3,\ldots,\beta_9) = &\beta_3 t + \beta_4 t^2 + \beta_5 t^3 + \beta_6 (t-70)^3_+ + \\ &\beta_7(t-140)^3_+ + \beta_8(t-220)^3_+ + \beta_9(t-300)^3_+ \end{aligned}$$

  where $(.)_+$ is the positive part of what is inside the parenthesis.

Although spline functions are not very commonly used in marketing, they have been used in modelling efforts in the past, for instance to help estimate irregular pricing effects (Kalyanam and Shively 1998) or to model reciprocation of punitive actions in channel relationships (Kumar et al. 1998). As Figure 2 shows, the use of splines in this case does allow for our model to follow rather well the camel-like distribution of flying behaviour among the top-tier frequent flyers over the first year of data, used to estimate our model.

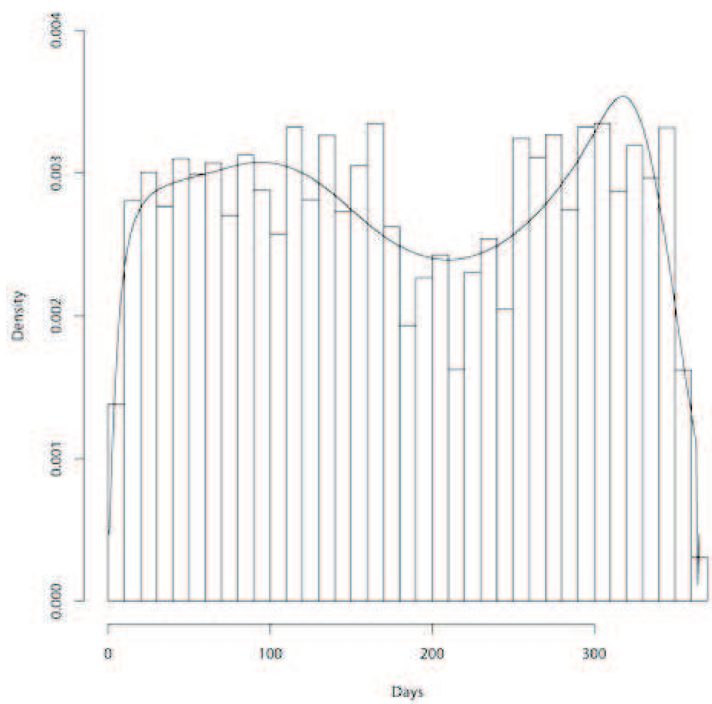

Figure 2: Adequacy of the nonhomogeneous process

5 Empirical Test of the Model

To assess the predictive quality of our model, we first explore its performance on predicting the number of flights to be flown by 5,000 members of a same cohort of top-tier frequent flyers within this loyalty program. For demonstration purposes, let us consider a scenario in which, as she prepares her marketing activities for the fall season, a customer relationship manager of this loyalty program is concerned about deploying extra effort to retain those Gold customers that are in danger of not qualifying for Gold status the next year. In order to target the right customers with a costly special offer, this manager wishes to target those with a moderate chance of actually qualifying for top-tier membership, between 50% and 90% as assessed using data available on August 1st of 2004. For each of these "gold" customers, our model returns the likelihood that these customers will remain top-tier members in 2006—that is, the likelihood that they will fly 20 flights or more during 2005. Those frequent flyers having already flown these 20 flights are assigned a probability of 100%. Table 2 shows these probabilities for 11 segments according to how likely they are to remain "gold" customers.

In addition, when summating the probabilities of all 3,438 customers who had not yet secured top-tier membership by August 1st of 2005, we get a prediction of 1,594 that do retain their "gold" status for 2006. In actuality 1,511 of them really do retain gold membership, that is, a 5% error.

Table 2: Model Fit According to Likelihood of Retaining Top-Tier Frequent Flyer Status

Probability intervals	Number of customers in interval (Aug. 1st, 2005)	Actual proportion of customers who retained top-tier membership by Dec. 31st, 2005
[0-10%[	1 149	3%
[10-20%[	250	16%
[20-30%]	179	30%
[30-40%[	129	33%
[40-50%[	134	49%
[50-60%[	160	50%
[60-70%[	150	56%
[70-80%[	147	69%
[80-90%[	205	72%
[90-100%[	935	92%
[100%] (already qualified by August 1st, 2004)	1 562	100%

To further assess the predictive performance of our model, we use the data available from January 1st, 2004 and August 1st, 2005 to extrapolate the rate function of our nonhomogeneous Poisson process between August 1st, 2005 and December 31th, 2005. As Figure 3 shows, our model allows rather precise prediction past the August 1st, 2005 date. This analysis demonstrates the high degree of validity of using nonhomogeneous mixed Poisson models for the purposes of forecasting a customer's future purchasing, conditional on his past buying behaviour and his activity to date.

Finally, as a last example of the usefulness of this approach to various instances, imagine our customer retention manager were interested in predicting likelihood of remaining top-tier customers at the beginning of each month. Let us consider the example of two Gold customers who both flew 25 flights over the first year. They both have an 83.5% likelihood of remaining Gold customers at the beginning of the year. Ultimately, Customer A will fly 23 qualifying flights this year thus conserving his top-tier status, whereas Customer B will fly only 18, meaning he will loose his top-tier status at the end of the year.

Figures 4 and 5 provide the 12 monthly 95% prediction intervals for Customers A and B. The dotted lines on both graphs indicate the total numbers of flights actually flown by the end of the year while the increasing solid curve represents the total number of flights taken at that point in time. As can be seen, and as an additional demonstration of the predictive ability of our model, the forecasted intervals always contain the actual, final number of lights taken for each of those two customers. Of course, the prediction interval also becomes smaller with time, as data accrue regarding both customers' actual behaviour.

On the basis of their respective flying activities, our model allows to estimate the probability that each of these two customers will take at least 20 flights at any point in time. For instance, a monthly review would provide the probabilities of taking at least 20 flights before the end of the year for each member (see Table 3).

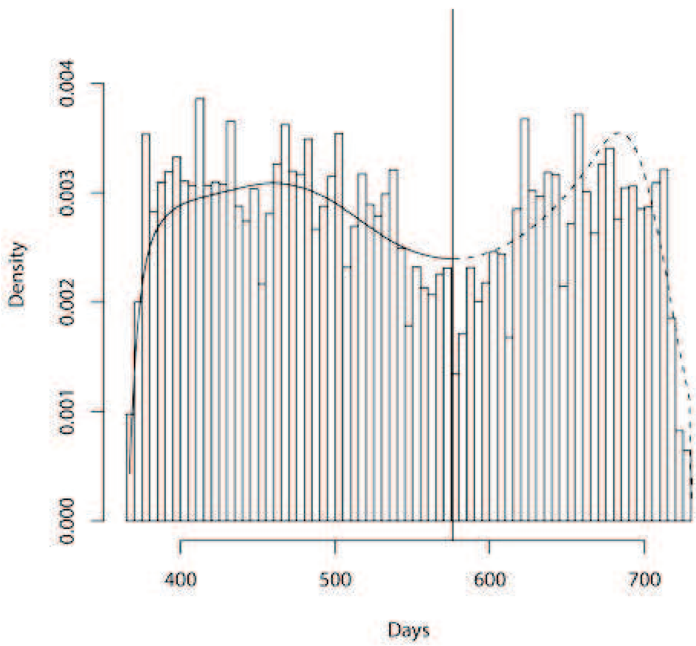

Figure 3: Accuracy of the forecasting

As can be seen in Table 3, while the probability of Customer A retaining his top-tier status by the end of the year remains high—above 60%—throughout the year, Customer B can be identified as a potentially his top-tier status as early as March or April. Considering that only 2 flights actually made the difference in the end, the airline company could have used such approaches as reminding Customer B of the value of his Gold membership to incentivize this customer into flying more in order to retain top-tier benefits into the next year. Adopting such "corrective" actions early on during the year would have likely left enough time for Customer B to better plan his flying activities for the remainder of the year.

6 Summary and Discussion

Retaining top-tier customers holds considerable importance for companies because of the net effect these customers have on any organization's bottom-line. This is why a large amount of marketing literature has focused on customer satisfaction with service quality and determinants of loyalty, among many others, as service quality is believed to be the key driver of top-tier customer retention. One key assumption in the customer lifetime value associated with top-tier customers is that they are not costly to market to since they do not need marketing—they are already acquired by the firm. However, in real life, even top-tier customers can leave the company for reasons other than service or product failures. For relationship managers, understanding which customers are likely to leave, and identifying which of these the organization still has a chance to retain as top-tier customers can be extremely

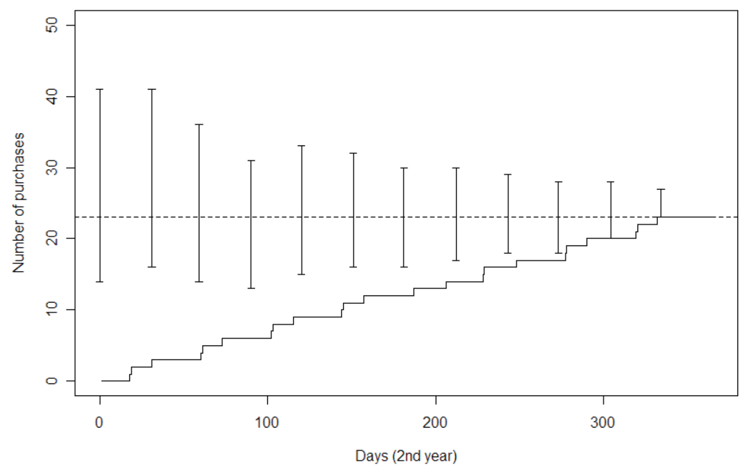

Figure 4: 95% prediction intervals for customer A

Table 3: Probabilities of taking 20 flights or more during this year, assessed at the beginning of each month based on historical data to date

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Customer A (23 flights actually taken at the end of the year)	.835	.899	.767	.608	.797	.834	.795	.842	.900	.907	1.000	1.000
Customer B (18 flights actually taken at the end of the year)	.835	.657	.442	.170	.441	.606	.535	.575	.636	.783	.413	.031

important. Indeed, managers will want to time their retention efforts, and target them precisely towards these customers they are likely to loose but may still retain provided the right actions are taken.

In this study, we have used nonhomogeneous Poisson processes with random effects to model top-tier customers' behaviour towards the firm by explicitly accounting for their expected consumption pattern over time. We estimated a model based on data from an airline's frequent flyer program where the behaviour is observed and stored for each customer. We then showed how well the model fit the reality after an initial data gathering period in the early phase of the year.

All prediction models are best applied in the specific situations where their critical assumptions are satisfied. Our approach is especially well suited in situations where a firm has a well-defined threshold (advertised or not) past which customers qualify for top-tier status and where this threshold must be met within a specific calendar time (although a slight

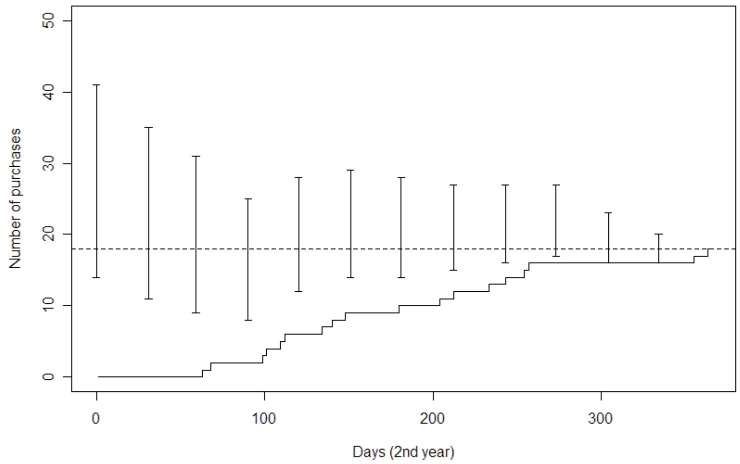

Figure 5: 95% prediction intervals for customer B

modification could be made to our model to allow this threshold to be met within a specific period (e.g., number of days), regardless of the calendar date). The use of spline functions also make our approach especially well suited for situations with irregular purchase behaviour, like seasonal or cyclical products or services. Aside from frequent traveler programs in airlines, hotels, and rental car companies, examples of such situations could be membership-based retailers (e.g., Costco, Sam's Club), credit cards, car repair shops, B2B services, and any type of business that actively collects information on its customers. In these circumstances, the potential drawback could of course be the availability of those appropriate measures. However, the proposed model needing as little as transaction dates per customers, it should be widely implementable in a large number of businesses.

Appendix A: Prediction of the total number of recurrent events amongst all the customers

For the prediction problem at hand, we can see from (4) that the predictive distribution of $\sum_{i=1}^{k} N_i(t_i, T)$ given $N(t)$ is a convolution (a sum) of $kNB\left(a + N_i\left(t_i\right), \left(b + F\left(t_i;\beta\right)\right) / \left(b + F\left(T;\beta\right)\right)\right)$ distributions with a probability function given by

$$
\begin{aligned}
P&\left[\sum_{i=1}^{k} N_i(t_i, T) = n | N(t); a, b, \beta\right] \\
&= \sum_{\{z_i : \sum_{i=1}^{k} z_i = n\}} P\left[N_i(t_i, T) = z_i | N_i(t_i); a, b, \beta\right] \\
&= \sum_{\{z_i : \sum_{i=1}^{k} z_i = n\}} \frac{\Gamma(a + N_i(t_i) + z_i)}{\Gamma(a + N_i(t_i)) z_i!} \left(\frac{F(T;\beta) - F(t_i;\beta)}{b + F(T;\beta)}\right)^{z_i} \left(\frac{b + F(t_i;\beta)}{b + F(T;\beta)}\right)^{a + N_i(t_i)} .
\end{aligned}
$$

A "plug-in" prediction interval is then obtained by finding the appropriate quantiles of this predictive distribution evaluated at $\hat{a}, \hat{b}$, and $\hat{\beta}$.

When n or k is large, it is more convenient to use simulations to approximate this predictive distribution. It is shown in Fredette and Lawless (2007) that a good approximation is given by

$$P\left[\sum_{i=1}^{k} N_i(t_i, T) = n | N(t); \hat{a}, \hat{b}, \hat{\beta}\right] \approx \sum_{j=1}^{M} \frac{\exp\{-u_j^*\}(u_j^*)^n}{n!}$$

Where u_j^* is obtained by simulating a convolution of k gamma random variables with parameters $\hat{a} + N_i(t_i)$ and $\frac{\hat{b}+F(t_i;\hat{\beta})}{F(T;\hat{\beta})-F(t_i;\hat{\beta})}$. In most applications, it is usually sufficient to use $M = 500$ or $M = 1,000$.

References

Batislam, Emine Persentili; Denizel, Meltem; Filiztekin, Alpay (2007), "Empirical validation and comparison of models for customer base analysis," *International Journal of Research in Marketing*, 24(3), 201–209.

Bhattacharya, C. B. (1998), "When Customers Are Members: Customer Retention in Paid Membership Contexts," *Journal of the Academy of Marketing Science*, 26(1), 31–44.

Bolton, Ruth 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.

Borle, Sharad; Singh, Siddharth S.; Jain, Dipak C. (2008), "Customer Lifetime Value Measurement," *Management Science*, 54(1), 100–112.

Buckinx, Wouter and Van Den Poel, Dirk (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.

Drozdenko, Ronald G. and Drake, Perry D. (2002), *Optimal Database Marketing*, Thousand Oaks, CA: Sage.

Fader, Peter S.; Hardie, Bruce G. S.; Ka Lok Lee (2005a), "RFM and CLV: Using Iso-Value Curves for Customer Base Analysis," *Journal of Marketing Research*, 42(4), 415–430.

Fader, Peter S.; Hardie, Bruce G. S.; Ka Lok Lee (2005b), "'Counting Your Customers' the Easy Way: An Alternative to the Pareto/NBD Model," *Marketing Science*, 24(2), 275–284.

Fredette, Marc; and Lawless, J.F. (2007), "Finite-Horizon Prediction of Recurrent Events, With Application to Forecasts of Warranty Claims," *Technometrics*, 49(1), 66–80.

Grandell, J. (1997), *Mixed Poisson Processes*, London: Chapman & Hall.

Gupta, Sunil and Donald R. Lehmann (2003), "Customer as assets," *Journal of Interactive Marketing*, 17(1), 9–24.

Harrel, F.E. (2001), *Regression Modeling Strategies*, New York: Springer.

Kalyanam, Kirthi; and Shively, Thomas S. (1998), "Estimating Irregular Pricing Effects: A Stochastic Spline Regression Approach," *Journal of Marketing Research*, 35(1), 16–29.

Kumar, Nirmalya; Scheer, Lisa K.; and Steenkamp, Jan-Benedict E. M. (1998), "Interdependence, Punitive Capability, and the Reciprocation of Punitive Actions in Channel Relationships," *Journal of Marketing Research*, 35(2), 225–235.

Lawless, J.F. (1987), "Regression Methods for Poisson Process Data," *Journal of the American Statistical Association*, 82, 808–815.

Lawless, J.F. and Fredette, Marc (2005), "Frequentist prediction intervals and predictive distributions," *Biometrika*, 92, 529–542.

Lewis, S. Michael (2005), "Research note: A dynamic programming approach to customer relationship pricing," *Management Science*, 56(6), 986–994.

Parasuraman, A., Zeithaml, V.A. and Berry, L.L. (1985), "A conceptual model of service quality and its implication," *Journal of Marketing*, 49 (Fall), 41–50.

Parasuraman, A., Zeithaml, V.A. and Berry, L.L. (1991), "Refinement and reassessment of the SERVQUAL scale," *Journal of Retailing*, 67, 420–450.

Pfeifer, Phillip E. and Carraway, Robert L. (2000), "Modeling Customer Relationships as Markov Chains," *Journal of Interactive Marketing*, 14(2), 43–55.

Reichheld, Frederick F. and Teal, Thomas (1996), *The Loyalty Effect: The Hidden Force Behind Growth, Profits, and Lasting Value*, Cambridge, MA: Harvard Business School Press.

Reichheld, Frederick F. (1993), "Loyalty-Based Management," *Harvard Business Review*, 71(2), 64–73.

Reichheld, Frederick.F. and Sasser, W.Earl (1990), "Zero Defections: Quality Comes to Services," *Harvard Business Review*, Sept-Oct, 301–307.

Ryals, Lynette J. and Rogers, Beth (2005), "Sales compensation plans—One size does not fit all," *Journal of Targeting, Measurement and Analysis for Marketing*, 13(4), 354–362.

Rust, Roland T., Lemon, Katherine N., and Zeithaml, Valarie A. (2004), "Return on Marketing: Using Customer Equity to Focus Marketing Strategy," *Journal of Marketing*, 68(1), 109–127.

Schweidel, David A; Fader, Peter S; Bradlow, Eric T. (2008), "Understanding Service Retention Within and Across Cohorts Using Limited Information," *Journal of Marketing*, 72(1), 82–94.

Simester, Duncan I.; Peng Sun; Tsitsiklis, John N. (2006), "Dynamic Catalog Mailing Policies," *Management Science*, 52 (5), 683–696.

Teunter, Ruud H.; and Klein Haneveld, Willem K. (2008), "Dynamic inventory rationing strategies for inventory systems with two demand classes, Poisson demand and backordering," *European Journal of Operational Research*, 190(1), 156–178.

Van Den Poel, Dirk and Larivière, Bart (2004), "Customer attrition analysis for financial services using proportional hazard models," *European Journal of Operational Research*, 157(1), 196–217.

Venkatesan, Rajkumar; Kumar, V; Bohling, Timothy (2007), "Optimal Customer Relationship Management Using Bayesian Decision Theory: An Application for Customer Selection," *Journal of Marketing Research*, 44(4), 579–594.

Wang, Hui-Ming; Kalwani, Manohar U.; and Akura, Tolga (2007), "A Bayesian multivariate Poisson regression model of cross-category store brand purchasing behaviour," *Journal of Retailing & Consumer Services*, 14(6), 369–382.