



Management Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Modeling Purchasing Behavior with Sudden "Death": A Flexible Customer Lifetime Model

Albert C. Bemmaor, Nicolas Gladys,

To cite this article:

Albert C. Bemmaor, Nicolas Gladys, (2012) Modeling Purchasing Behavior with Sudden "Death": A Flexible Customer Lifetime Model. Management Science 58(5):1012-1021. <http://dx.doi.org/10.1287/mnsc.1110.1461>

Full terms and conditions of use: <http://pubsonline.informs.org/page/terms-and-conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2012, INFORMS

Please scroll down for article—it is on subsequent pages



INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Modeling Purchasing Behavior with Sudden “Death”: A Flexible Customer Lifetime Model

Albert C. Bemmaor, Nicolas Gladly

ESSEC Business School, 95021 Cergy-Pontoise Cedex, France
{bemmaor@essec.edu, gladly@essec.edu}

This study proposes a new customer lifetime model: the gamma/Gompertz distribution (G/G). The advantage of this model relative to the well-known Pareto distribution is twofold: (i) its probability density function can exhibit a mode at zero or an interior mode, and (ii) it can be skewed to the right or to the left. We combine the G/G with a negative binomial distribution (NBD) and obtain the moments of the distribution of the number of transactions over $(0, T]$ and $(T, T + T^*]$. Out of six data sets, the G/G/NBD model provides a notable improvement in the log-likelihood over the Pareto/NBD model in four data sets. It can indicate substantial differences in expected residual lifetimes compared to the Pareto/NBD and induce a retention rather than acquisition policy. On the average, the G/G/NBD exhibits slightly better forecasts of the mean number of transactions than the Pareto/NBD.

Key words: buyer behavior; mixture models; catalog retailing; Gompertz distribution

History: Received June 28, 2011; accepted September 7, 2011, by Pradeep Chintagunta, marketing. Published online in *Articles in Advance* December 2, 2011.

Introduction

The purpose of this paper is to propose a new customer lifetime model for implementation in the noncontractual setting of customer-base analysis, i.e., when defection times are unknown to the firm. In an influential study, Schmittlein et al. (1987) combine the negative binomial distribution (NBD) with a Pareto model to capture attrition. The Pareto/NBD has led to several applications and extensions (see, e.g., Fader et al. 2005; Ho et al. 2006; Krafft 2002; Reinartz and Kumar 2000, 2003). After a series of papers that showed the adequacy of the Pareto/NBD model, recent papers have pointed to its limitations. Using transactions data from a membership-based direct marketing company, Borle et al. (2008, Figure 3) show that the customer lifetime curve exhibits a peak at around 60 weeks after the initial transaction. In a study on the donations to a nonprofit organization over a six-year period, Fader et al. (2010) find that the Pareto/NBD (i) underestimates the proportion of repeat donors and (ii) underpredicts the rate of attrition of the number of repeat donations. Abe (2009) uses Markov chain Monte Carlo (MCMC) methods to estimate individual-specific parameters of the transaction model for “one-to-one” marketing purposes. He keeps the exponential distribution as an individual-level model of customer lifetime but replaces the gamma distribution with a lognormal distribution. Like the Pareto distribution, the resulting probability density function exhibits a mode at zero that implies that a randomly picked customer is most

likely to churn immediately after the initial transaction. Singh et al. (2009, p. 181) propose an improved data augmentation estimation method that reportedly can accommodate “any of the commonly available statistical distributions.” Rather than tinkering with a variety of possible probability distributions that can lead to an overly complex model, the analyst may prefer to apply a distribution that is relatively easy to use. The potential deficiencies of the Pareto/NBD call for a change either in the focus variable or in the model formulation. Whereas Fader et al. (2010) select the former option, we opt for the latter alternative and propose a flexible model of customer lifetime. The main reason for focusing on the customer lifetime process is that the data requirements for parameter estimation (i.e., recency and frequency) remain the same as those for the Pareto/NBD; hence, the combined models are directly comparable in terms of implementation, predictions, and fit.

This study proposes a new customer lifetime model that is a gamma mixing of Gompertz distributions (G/G). Compared to the Pareto distribution, its advantage is twofold:

- (i) its probability density function (p.d.f.) can exhibit a mode at zero or an interior mode;
- (ii) its p.d.f. can be skewed to the right or to the left.

A shift of the mode away from zero can occur when (i) the offering is strongly differentiated and (ii) the organization benefits from a strong reputation. Charities, specialty stores, reputed hard discounters,

upscale catalog retailers, and high-end hotels can provide such examples. The Gompertz distribution is a natural conjugate to a gamma distribution. When combined with the NBD, the G/G distribution permits to forecast the number of future transactions of a new customer from his or her purchasing history. Besides, in contrast with the Pareto distribution, the moments of the G/G distribution are finite for all the values of the parameters. Finally, the G/G distribution shows that a Pareto-like decreasing churn (death) rate at the aggregate level can be consistent with an increasing churn rate at the individual level. This pattern can be explained by a sorting effect in a heterogeneous population (i.e., the individuals with the higher propensities to search tend to "die" earlier than the others).

The Model

The Individual-Level Model

The Customer Lifetime Model. The cumulative distribution function (c.d.f.) of a Gompertz distribution is such as (Gompertz 1825; Johnson et al. 1995, pp. 82–85; Sheikh et al. 1989)

$$F(\tau | \eta) = 1 - \exp(-\eta(e^{b\tau} - 1)), \quad \eta, b > 0, \tau > 0. \quad (1)$$

As the probability that a customer dies before a fixed time τ (Equation (1)) increases with η , the shape parameter η can be interpreted as a customer's propensity to search for alternatives. The parameter b is a scale parameter. The p.d.f. is such as

$$f(\tau | \eta) = b\eta e^{b\tau} e^{-\eta} \exp(-\eta e^{b\tau}). \quad (2)$$

The mode of the p.d.f. is as follows:

$$\begin{aligned} \tau_{\eta}^* &= (1/b) \ln(1/\eta), \quad \text{with } 0 < F(\tau_{\eta}^*) < 1 - e^{-1} \\ &= 0.632121, \quad 0 < \eta < 1, \\ \tau_{\eta}^* &= 0, \quad \eta \geq 1. \end{aligned} \quad (3)$$

The Gompertz distribution can take a large variety of shapes. It can be skewed to the right or to the left. Figure 1 depicts the shapes of the c.d.f. and p.d.f. The expected customer's lifetime is such as

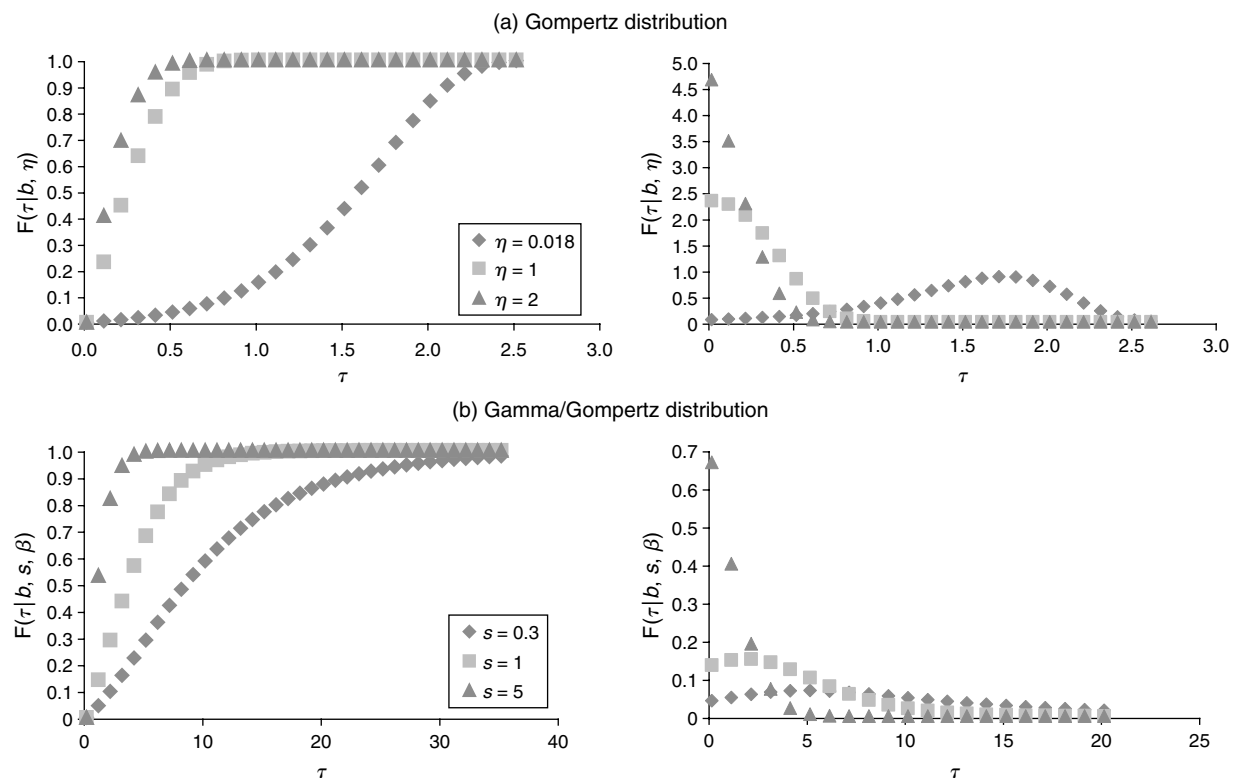
$$\mathbb{E}(\tau | \eta) = -(1/b)e^{\eta} \text{Ei}(-\eta), \quad (4)$$

where $\text{Ei}()$ denotes the exponential integral (Abramowitz and Stegun 1972, p. 228):

$$\text{Ei}(x) = - \int_{-x}^{\infty} (e^{-u}/u) du. \quad (5)$$

For given b , when η gets close to zero, Equation (4) approaches ∞ . Let the origin of the time scale be the

Figure 1 Gompertz and Gamma/Gompertz Distributions



Notes. Each panel depicts the cumulative distribution function on the left-hand side and the probability density function on the right-hand side. The parameters are as follows: $b = 2.322$ and $\eta = 0.018, 1$, and 2 (panel (a)); $b = 0.4$, $\beta = 3$, and $s = 0.3, 1$, and 5 (panel (b)).

time of the first transaction and let θ be the unobserved lifetime of customer i over the observation period $(0, T_i]$. It can be defined as

$$\theta = \min(\tau, T_i), \quad 0 < \theta < T_i. \quad (6)$$

The expected value of θ is such as

$$\mathbb{E}(\theta | \eta, T_i) = (1/b)e^\eta [\text{Ei}(-\eta e^{bT_i}) - \text{Ei}(-\eta)]. \quad (7)$$

For fixed values of b and T_i , $\mathbb{E}(\theta | \eta, T_i)$ decreases monotonically as η increases. When T_i becomes large, (7) approaches (4). The second moment of θ about zero is such as

$$\begin{aligned} \mathbb{E}(\theta^2 | \eta, T_i) &= \exp(-\eta(e^{bT_i} - 1))T_i^2 \\ &+ \eta b e^\eta \int_0^{T_i} \tau^2 e^{b\tau} \exp(-\eta e^{b\tau}) d\tau. \end{aligned} \quad (8)$$

The variance is given by

$$\text{Var}(\theta | \eta, T_i) = \mathbb{E}(\theta^2 | \eta, T_i) - \mathbb{E}(\theta | \eta, T_i)^2. \quad (9)$$

For fixed values of b and T_i , $\text{Var}(\theta | \eta, T_i)$ is a non-monotone function of η : it increases to a maximum and decreases. The coefficient of variation (standard deviation/mean) approaches one from below as η increases.

The hazard rate $z(\tau | \eta) = f(\tau | \eta)/(1 - F(\tau | \eta))$ is an exponentially increasing function of the time elapsed since "birth" (i.e., the date of the first transaction ever). As a customer stays longer with a firm, his or her likelihood to churn (or to die) at any instant given that he or she has not churned yet, increases (positive duration dependence). The rate of increase is proportional to the propensity to search η . The pattern of the hazard rate is compatible with the results obtained by Jamal and Bucklin (2006, Table 5) on a direct-to-home satellite TV provider: the exponent parameter of an individual-level lifetime Weibull distribution is larger than one for all three segments of customers.

Combining the Lifetime and Transaction Processes. Let x_i be the number of repeat transactions made by customer i over the interval $(0, T_i]$. When the transaction process is consistent with a Poisson distribution with parameter λ while a customer is active, the resulting mean number of repeat transactions over the interval $(0, T_i]$ is such as

$$\mathbb{E}(x_i | \lambda, \eta, T_i) = \lambda \mathbb{E}(\theta | \eta, T_i). \quad (10)$$

The variance of x_i is given by

$$\begin{aligned} \text{Var}(x_i | \lambda, \eta, T_i) \\ = \mathbb{E}(x_i | \lambda, \eta, T_i) + \lambda^2 \text{Var}(\theta | \eta, T_i); \end{aligned} \quad (11)$$

hence, $\text{Var}(x_i | \lambda, \eta, T_i)$ is larger than $\mathbb{E}(x_i | \lambda, \eta, T_i)$.

The Heterogeneity Assumptions

We assume the following:

(i) The parameter η of the Gompertz distribution of a customer's lifetime is distributed gamma with shape parameter s and scale parameter β (mean = s/β). The coefficient of variation is equal to $s^{-1/2}$: the smaller s is, the stronger the heterogeneity of customer lifetimes.

(ii) The scale parameter b of the Gompertz distribution is constant across customers.

(iii) The mean purchasing frequency λ over a unit time period while a customer is active ("alive") is distributed gamma with shape parameter r and scale parameter α (mean = r/α) across the population.

(iv) λ and η are independent.

The gamma distribution is a natural conjugate prior to the Gompertz distribution. Let T_i be the length of customer i 's tenure with the firm. When the prior distribution of the propensity to search η is gamma with parameters s and β and the p.d.f. of $\tau | \eta$ is Gompertz, the distribution of $\eta | \tau > T_i$ is gamma with parameters s and $\beta^* = \beta + e^{bT_i} - 1$. The posterior mean of η is equal to s/β^* . The coefficient of variation of $\eta | \tau > T_i$ across the population remains unchanged: it is equal to $s^{-1/2}$.

The Aggregate-Level Model

The Customer Lifetime Model. The c.d.f. of the G/G is given by

$$\begin{aligned} F(\tau) &= 1 - \beta^s / (\beta - 1 + e^{b\tau})^s, \quad \tau > 0, b, s, \beta > 0, \\ &= 1 - e^{-bs\tau}, \quad \beta = 1. \end{aligned} \quad (12)$$

When β is equal to one, the G/G reduces to an exponential distribution. The p.d.f. is such as

$$\begin{aligned} f(\tau) &= b s e^{b\tau} \beta^s / (\beta - 1 + e^{b\tau})^{s+1}, \quad b, s, \beta > 0, \\ &= b s e^{-bs\tau}, \quad \beta = 1. \end{aligned} \quad (13)$$

The mean customer lifetime is such as

$$\begin{aligned} \mathbb{E}(\tau | b, s, \beta) \\ &= (1/b)(1/s) {}_2F_1(s, 1; s+1; (\beta-1)/\beta), \quad b, s > 0, \beta \neq 1, \\ &= (1/b)[\beta/(\beta-1)] \ln(\beta), \quad b > 0, s=1, \beta \neq 1, \\ &= 1/(bs), \quad b, s > 0, \beta = 1, \end{aligned} \quad (14)$$

where ${}_2F_1(a, b; c; z)$ denotes a Gauss hypergeometric function (Abramowitz and Stegun 1972, p. 558). To obtain the conditional mean residual lifetime given customer i is alive at time T_i , we replace β by β^* in Equation (14). The conditional mean residual lifetime increases monotonically toward the following asymptote:

$$(1/b)(1/s) {}_2F_1(s, 1; s+1, 1). \quad (15)$$

For given b , Equation (15) is all the closer to zero as the customer lifetime distribution approaches complete homogeneity (i.e., s is large). Depending on the value of the asymptote, the G/G can capture an apparently bounded inertia. The moments (about zero) of the lifetime distribution of a randomly picked customer i over the interval $(0, T_i]$ are such as

$$\begin{aligned} \mathbb{E}(\theta^k | b, s, \beta, T_i) \\ = [\beta / (\beta + e^{bT_i} - 1)]^s T_i^k \\ + bs\beta^s \int_0^{T_i} \tau^k e^{b\tau} (\beta + e^{b\tau} - 1)^{-(s+1)} d\tau, \quad k=1, 2, \dots \end{aligned} \quad (16)$$

The mode of the p.d.f. is given by

$$\begin{aligned} \tau^* = 0, \quad 0 < \beta \leq s + 1, \\ \tau^* = (1/b) \ln[(\beta - 1)/s], \quad \text{with } 0 < F(\tau^*) < 0.632121, \\ \beta > s + 1. \end{aligned} \quad (17)$$

The G/G distribution is a flexible model of customer lifetime: its p.d.f. can be skewed to the right or to the left (see Figure 1). As a randomly picked customer's tenure with the firm increases, the model predicts that his or her likelihood to die (i) decreases toward bs when β is less than one, (ii) remains constant (and equal to bs) when β is equal to one, and (iii) increases toward bs when β is larger than one. The model can accommodate a variety of possible scenarios.

Combining the G/G with the NBD. The expected number of repeat transactions for a randomly picked customer i over a fixed interval $(0, T_i]$ is such as

$$\mathbb{E}(x_i | r, \alpha, b, s, \beta, T_i) = (r/\alpha) \mathbb{E}(\theta | b, s, \beta, T_i). \quad (18)$$

The variance is such as

$$\begin{aligned} \text{Var}(x_i | r, \alpha, b, s, \beta, T_i) \\ = \mathbb{E}(x_i | r, \alpha, b, s, \beta, T_i) + [r(r+1)/\alpha^2] \mathbb{E}(\theta^2 | b, s, \beta, T_i) \\ - \mathbb{E}^2(x_i | r, \alpha, b, s, \beta, T_i). \end{aligned} \quad (19)$$

Let t_{xi} be the time elapsed until customer i 's last transaction during the observation period $(0, T_i]$. The conditional expectation of the number of future transactions x_i^* over the period $(T_i, T_i + T^*]$ for a randomly picked customer i with vector (x_i, t_{xi}, T_i) is given by

$$\begin{aligned} \mathbb{E}(x_i^* | r, \alpha, b, s, \beta, x_i, t_{xi}, T_i, T^*) \\ = \mathbb{P}(\tau > T_i | r, \alpha, b, s, \beta, x_i, t_{xi}, T_i) \\ \cdot \mathbb{E}(x_i^* | r^*, \alpha^*, b, s, \beta^*, T^*), \end{aligned} \quad (20)$$

where $\mathbb{E}(x_i^* | r^*, \alpha^*, b, s, \beta^*, T^*)$ is shown in (18) with $r^* = r + x_i$, $\alpha^* = \alpha + T_i$, β^* , and T^* instead of r , α , β , and T_i , respectively, and $\mathbb{P}(\tau > T_i | r, \alpha, b, s, \beta, x_i, t_{xi}, T_i)$ is given in Equation (29). We can obtain the expected residual number of transactions by replacing $\mathbb{E}(\theta^* | b, s, \beta^*, T^*)$ in Equation (20)

by the conditional mean residual lifetime given customer i is alive at time T_i . The conditional variance of the number of future transactions over the period $(T_i, T_i + T^*]$ is such as

$$\begin{aligned} \text{Var}(x_i^* | r, \alpha, b, s, \beta, x_i, t_{xi}, T_i, T^*) \\ = \mathbb{P}(\tau > T_i | r, \alpha, b, s, \beta, x_i, t_{xi}, T_i) \\ \cdot \mathbb{E}(x_i^{*2} | r^*, \alpha^*, b, s, \beta^*, T^*) \\ - \mathbb{E}^2(x_i^* | r, \alpha, b, s, \beta, x_i, t_{xi}, T_i, T^*), \end{aligned} \quad (21)$$

where $\mathbb{E}(x_i^{*2} | r^*, \alpha^*, b, s, \beta^*, T^*)$ is obtained by summing Equation (19) and the square of Equation (18) with r^* , α^* , β^* , and T^* instead of r , α , β , and T_i , respectively. The mean and variance of the number of future transactions x_i^* for a random picked customer i over the period $(T, T + T^*]$ with $T \geq \max(T_i)$ are such as

$$\begin{aligned} \mathbb{E}(x_i^* | r, \alpha, b, s, \beta, T, T^*) \\ = \mathbb{E}[\mathbb{E}(x_i^* | r, \alpha, b, s, \beta, x_i, t_{xi}, T_i, T^*)], \end{aligned} \quad (22)$$

$$\begin{aligned} \text{Var}(x_i^* | r, \alpha, b, s, \beta, T, T^*) \\ = \text{Var}[\mathbb{E}(x_i^* | r, \alpha, b, s, \beta, x_i, t_{xi}, T_i, T^*)] \\ + \mathbb{E}[\text{Var}(x_i^* | r, \alpha, b, s, \beta, x_i, t_{xi}, T_i, T^*)]. \end{aligned} \quad (23)$$

We can replace Equations (20) and (21) by Equations (18) and (19), respectively, to obtain the mean and variance of the number of transactions over the period $[0, T]$.

Overall, the G/G can be considered as a generalization of the Pareto distribution as it embeds the shapes of its p.d.f. and those of its hazard rate as special cases. It includes several appealing properties, such as (i) a flexible probability density function that can be skewed to the right or to the left, (ii) natural conjugacy, (iii) a relatively small number of parameters, and (iv) a monotonically increasing or decreasing or constant hazard rate. It predicts an increasing mean residual lifetime to an asymptote (bounded inertia). We derive the first two moments of the G/G/NBD model to obtain the predictions of future behavior.

Empirical Application and Simulation

The Data

We use six data sets (Table 1). The U.S. charity, the reproduction gift catalog company, the nonprofit organization, and the specialty catalog company appear to provide differentiated offerings.

The focus variable x_i corresponds to the number of repeat donations or the number of repeat orders placed by customer i in a predetermined time interval. The Direct Marketing Educational Foundation (DMEF) provided all the data sets, except the CDNOW data (Abe 2009, Fader et al. 2005). It used the U.S. charity data in the context of a competition in the academic community (Malthouse 2009).

Table 1 Data Description

Data set	Business description	Observation period for the data set	Number of supporters in the data set	Cohort size	Length of the observation period for the cohort	Time split (estimation/validation) ^a	Number of transactions in estimation period	Number of transactions in validation period	Average T_i
CDNOW	Electronic commerce (music CDs)	January 1, 1997–June 30, 1998. The customers made their first purchase (ever) at CDNOW during the first quarter of 1997.	2,357	2,357	78 weeks	39/39	2,457	1,882	32.716
U.S. charity	Leading U.S. charity	January 2, 2002–August 31, 2008. The supporters made their first donation during the first half of 2002.	21,166	21,166	6.69 years	1/–; 3.54/1.15; 3.54/3.15; 4.69/2	5,280; 26,341; 32,832	–; 6,491; 16,964 ^b ; 10,477	0.796; 3.30; 4.41
Reproduction gift catalog company	Upscale gift business that mails general and specialized catalogs	December 1971–December 1992	101,532	11,399 10,330	3.33 years 6.33 years	3/0.33 6/0.33	4,377 9,964	802 765	2.34 5.333
Nonprofit organization	Uses direct mail to solicit contributions	October 1986–December 1995	99,200	7,385 3,605	1,092 days 1,392 days	1,000/92 1,300/92	69,475 53,973	2,426 1,358	833.32 1,174.50
Catalog company	Business with multiple divisions each mailing different catalogs	June 1972–December 1995	96,551	11,672 3,125	2.5 years 7.5 years	2/0.5 7/0.5	10,603 24,868	223 128	1.746 6.436
Specialty catalog company	Long-time specialty company that mails full line and seasonal catalogs	January 1971–December 1995	106,284	7,947 5,120	6.33 years 15.33 years	6/0.33 15/0.33	12,550 55,128	317 278	5.802 14.573

^aThe time units are given in the preceding column.^bThe first two numbers are to be compared with 26,341 transactions in the estimation period. There is no validation period with the first estimation period.

Estimation

Let N be the number of customers in the sample. We use the maximum likelihood method that consists of maximizing the following function:

$$LL = \sum_{i=1}^N \ln L(r, \alpha, b, s, \beta | x_i, t_{xi}, T_i), \quad (24)$$

where $L(r, \alpha, b, s, \beta | x_i, t_{xi}, T_i)$ is given in (28). The integrals in (24) can be evaluated numerically. (We use `quadgk` in MATLAB for the computation of the integrals and `fmincon` for the maximization of the log-likelihood. For example, it takes eight minutes to estimate the parameters of the G/G/NBD model on the CDNOW data set with a standard desktop personal computer.) When s approaches ∞ , the G/G/NBD becomes the Gompertz/NBD. For the Pareto/NBD, the notations are the same as those in Schmittlein et al. (1987): s and β are the parameters of the Pareto distribution (mean = $\beta/(s - 1)$, $s > 1$). The likelihood function is given in Fader et al. (2005, Equation (A1)). When s approaches ∞ , the Pareto distribution becomes an exponential distribution. When s gets close to zero, the G/G/NBD and the Pareto/NBD reduce to the NBD.

Empirical Results

We report the parameter estimates in Table 2. The full four-parameter Pareto/NBD model is estimable in two data sets out of six, i.e., when the mode of the customer lifetime distribution is zero. The CDNOW data is one of those two cases. For this data set, (i) the maxima of the log-likelihoods for the G/G/NBD and Pareto/NBD are the same; (ii) the aggregate-level hazard rate of the G/G is monotonically declining; (iii) the asymptote to the mean residual lifetime is 1,254 years; and (iv) for both models, the expected number of active customers at the end of the first 39 weeks is equal to 1,052. Both models look observationally equivalent. The G/G model predicts (i) a mean residual lifetime of 7.71 years across customers at the end of the 39-week observation period to be compared with 8.18 years at birth, i.e., a 5.7% increase; and (ii) a mean residual number of transactions of 19.60 versus 22.25 at birth, i.e., an 11.9% decrease. The correlation between the number of transactions in the initial 39-week period and the expected residual number of transactions equals 0.856, i.e., a rather large number given the possible occurrence of sudden "death." In the catalog company data (two-year estimation period), the hazard rate declines monotonically but the asymptote to the conditional mean residual lifetime given a customer is alive at the end of the estimation period, is 1.76 years. (The conditional median residual lifetime is 4.36 years for the Pareto/NBD.) In the remaining four data sets, nested versions of

the Pareto/NBD apply: the Pareto/NBD reduces to an Exponential/NBD or, as it happens with the U.S. charity data (one-year estimation period), it becomes a simple NBD, like the G/G/NBD does. In total, Table 2 shows the estimation of five models. The simple NBD model with no death occurs when the observation period is too short for parameter identification.

Table 3 compares the fit of the combined models. The Pareto/NBD provides a better fit than the G/G/NBD in two markets: CDNOW and catalog business (two-year estimation period), i.e., two relatively standard services. In the other data sets, the G/G/NBD leads to substantial gains in the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) with one additional parameter. (Note that the G/G/NBD outperforms the Pareto/NBD for the catalog company data also when the length of the estimation period is seven years.) In two markets (reproduction gift catalog company and specialty catalog company), the customer lifetime distribution is skewed to the left. The majority of the customers tend to stay for a relatively long time with the firm: i.e., more than 3.6 years (mean) for the reproduction gift catalog company (six-year estimation period) and more than 4.2 years for the specialty catalog company (six-year estimation period).

To assess the descriptive and forecasting accuracy of the model, we derive the following predictions: (i) the mean number of transactions during the estimation period $(0, T]$ and the subsequent period $(T, T + T^*]$ (Equation (22)), and (ii) the standard deviation of the number of transactions during each period (Equation (23)). Table 4 reports the descriptive and forecasting accuracy of the models. Across the six data sets, the fit of the G/G/NBD to the sample mean number of transactions is notably better than that of the Pareto/NBD with a 79% decrease in the mean absolute deviation (0.015 versus 0.071) and an 87% decrease in the root mean squared error (0.020 versus 0.155). The fit to the sample standard deviation is slightly better for the Pareto/NBD with 14% and 4% improvements on both criteria, respectively (0.22 versus 0.18 and 0.42 versus 0.18). On the average, the G/G/NBD tends to provide slightly more accurate forecasts of the mean number of transactions than the Pareto/NBD with a 15.2% improvement in the mean absolute deviation (0.195 versus 0.230) and a 17.0% improvement in the root mean squared error (0.289 versus 0.348). Perhaps most important, both models can differ in their substantive implications: For example, for the U.S. charity data, they lead to essentially the same forecast of the mean number of donations during the subsequent two years after the initial 4.69 years (0.440 versus 0.437), but the Pareto/NBD implies complete homogeneity of customer lifetimes whereas the G/G/NBD depicts a relatively strong heterogeneity ($s = 0.378$). According to

Table 2 Maximum Likelihood Parameter Estimates

Data set	Sample size	Estimation period	Pareto/NBD					G/G/NBD							
			r	α	s	s/β	β	LL	r	α	b	s	s/β	β	LL
CDNOW	2,357	39 weeks	0.553 (0.046) ^{1a}	10.578 (0.84)	0.606 (0.075)		11.668 (2.65)	−9,594.98	0.553 (0.049)	10.578 (0.949)	0.0002 (0.0000)	0.603 (0.039)	0.0026 (0.0004)	—	−9,594.98
U.S. charity	21,166	1 year ^b	0.504 (0.015)	1.601 (0.047)	0		—	−10,565.60	0.504 (0.015)	1.601 (0.047)	—	0	—	—	−10,565.10
U.S. charity	21,166	3.54 ^c years	0.62 (0.009)	1.471 (0.0077)		0.068 (0.0047)	−43,119.26		0.563 (0.000)	1.416 (0.0002)	4.106 (0.0002)		0.000001 (0.0000001)		−42,981.85
U.S. charity	21,166	4.69 years	0.632 (0.015)	1.384 (0.034)		0.136 (0.004)	−52,551.96		0.558 (0.016)	1.291 (0.015)	0.874 (0.204)	0.378 (0.131)	9.040 (2.404)		−52,471.07
Reproduction gift catalog company	11,399	3 years	3.015 (0.363)	9.167 (1.051)		0.688 (0.021)	−11,497.03		1.258 (0.060)	4.940 (0.252)	2.322 (0.091)		0.0180 (0.003)		−11,243.59
Reproduction gift catalog company	10,330	6 years	2.109 (0.129)	6.502 (0.380)		0.249 (0.006)	−24,672.65		1.205 (6 · 10 ^{−6})	4.430 (0.0002)	0.750 (4 · 10 ^{−5})	90.807 (10 ^{−6})	2.038.0 (0.0004)		−24,453.65
Nonprofit organization	7,385	1,000 days	7.806 (0.255)	680.09 (22.73)		5.0 · 10 ^{−5} (0.5 · 10 ^{−5})	−379,143.94		7.764 (0.132)	673.03 (11.01)	0.00331 (0.0003)	0.345 (0.048)	65.866 (6.936)		−379,118.07
Nonprofit organization	3,605	1,300 days	10.30 (0.082)	802.95 (0.665)		1.6 · 10 ^{−5} (0.3 · 10 ^{−5})	−288,368.20		10.461 (0.100)	810.21 (7.220)	0.0095 (0.0008)	0.0147 (0.005)	1,039.87 (6.693)		−288,330.54
Catalog company	11,672	2 years	1.627 (0.052)	1.352 (0.054)	0.646 (0.063)		0.268 (0.055)	−13,596.91	1.642 (0.118)	1.357 (0.086)	0.482 (0.374)	0.477 (0.093)	0.0866 (0.050)		−13,596.13
Catalog company	3,125	7 years	1.864 (0.0554)	1.490 (0.050)		0.00330 (0.0008)	−14,884.55		1.872 (0.053)	1.429 (0.045)	1.613 (0.082)	0.207 (0.042)	6,000.0 (1,717.9)		−14,622.35
Specialty catalog company	7,947	6 years	1.448 (0.075)	2.705 (0.132)		0.289 (0.006)	−23,699.42		0.673 (0.038)	1.803 (0.013)	3.132 (0.305)		10 ^{−6} (10 ^{−6})		−22,556.14
Specialty catalog company	5,120	15 years	2.327 (0.054)	2.918 (0.075)		0.0174 (0.0008)	−64,462.02		2.506 (0.355)	2.915 (0.0012)	1.055 (0.355)		10 ^{−6} (0.5 · 10 ^{−5})		−60,241.19

^aStandard errors are shown in parentheses.^bBoth models reduce to the NBD.^cFor these data, the Pareto and the G/G distributions reduce to an exponential distribution and to a Gompertz distribution, respectively.

Table 3 Fit of the Alternative Models and Characteristics of the Superior Model

Data/estimation period	AIC ^a		BIC ^b		Characteristics of the superior model	
	Pareto /NBD	G/G/NBD	Pareto/NBD	G/G/NBD	Mode ^c	Skewness $F(\tau^*)^d$
CDNOW						
39 weeks	19,197.96	19,199.96	19,221.02	19,228.79	0	0
U.S. charity ^e						
3.54 years	86,244.52	85,971.70	86,268.40	86,003.54	3.36	0.63
4.69 years	105,109.92	104,952.14	105,133.80	104,991.94	3.5	0.36
Reproduction gift catalog company						
3 years	23,000.06	22,497.18	23,022.08	22,533.89	1.7	0.63
6 years	49,351.30	48,917.30	49,373.03	48,953.51	4.1	0.61
Nonprofit organization						
1,000 days	758,293.88	758,246.14	758,314.60	758,280.68	1,582.0	0.37
1,300 days	576,742.40	576,671.08	576,760.97	576,702.03	1,313.6	0.06
Catalog company						
2 years	27,201.82	27,202.28	27,221.91	27,226.44	0	0
7 years	29,775.10	29,254.70	29,793.24	29,284.94	6.4	0.31
Specialty catalog company						
6 years	47,404.84	45,120.28	47,425.78	45,148.20	4.4	0.63
15 years	128,930.04	120,490.38	128,949.66	120,516.54	13.1	0.63

^aAIC = $-2LL + 2p$, where p is the number of parameters. The lower AIC is, the better the fit.

^bBIC = $-2LL + \ln(N)p$, where N is the sample size.

^cThe time units are shown in the first column.

^dWe report the ordinate of the c.d.f. at the mode. When $F(\tau^*)$ is larger than 0.5, the customer lifetime distribution is skewed to the left. Otherwise, it is skewed to the right.

^eWe omit the case where the Pareto/NBD and G/G/NBD reduce to the NBD model.

Table 4 Descriptive and Forecasting Accuracy of the Models

Data	Estimation and validation periods ^a	Estimation sample						Validation sample					
		Mean			Standard deviation			Mean			Standard deviation		
		Actual	P/NBD	G/G/NBD	Actual	P/NBD	G/G/NBD	Actual	P/NBD	G/G/NBD	Actual	P/NBD	G/G/NBD
CDNOW	39, 39 weeks	1.042	1.071	1.071	2.190	2.104	2.104	0.798	0.707	0.707	2.057	2.011	2.012
U.S. charity ^b	3.54, 1.15 years	1.244	1.246	1.244	2.299	2.015	2.005	0.307	0.373	0.250	0.913	1.132	0.724
U.S. charity	3.54, 3.15 years	1.244	1.246	1.244	2.299	2.015	2.005	0.801	0.955	0.664	2.242	2.079	1.617
U.S. charity	4.69, 2 years	1.551	1.53	1.538	2.940	2.564	2.592	0.495	0.440	0.437	1.512	1.183	0.935
Reproduction gift catalog company	3 years, 4 months	0.384	0.381	0.385	0.723	0.725	0.727	0.0704	0.0186	0.0245	0.312	0.144	0.163
Reproduction gift catalog company	6 years, 4 months	0.965	0.957	0.967	1.402	1.399	1.406	0.0741	0.0259	0.0191 ^c	0.302	0.171	0.144
Nonprofit organization	1,000 days, 92 days	9.408	9.365	9.396	4.953	4.860	4.842	0.329	1.010	0.985	0.472	1.046	1.238
Nonprofit organization	1,300 days, 92 days	14.972	14.924	14.953	6.223	6.258	6.202	0.377	1.157	0.988	0.488	0.595	1.045
Catalog company	2 years, 6 months	0.908	0.948	0.950	1.546	1.619	1.621	0.0191	0.141	0.142	0.153	0.331	0.503
Catalog company	7 years, 6 months	7.958	7.967	7.986	7.847	6.549	6.585	0.0410	0.596	0.405	0.220	0.448	0.827
Specialty catalog company	6 years; 4 months	1.579	1.505	1.578	2.266	2.406	2.499	0.0704	0.0211	0.00322	0.312	0.160	0.165
Specialty catalog company	15 years, 4 months	10.767	10.264	10.780	8.035	7.965	7.649	0.0741	0.175	0.00315	0.302	0.450	0.164
Overall mean		4.616	4.560	4.623	3.675	3.497	3.476	0.288	0.468	0.386	0.774	0.812	0.795

^aThe first number denotes the length of the estimation period.

^bWe derive two sets of forecasts with the same parameter estimates. We omit one row in the computation of the overall mean over the estimation sample.

^cFor this data set, the parameter s of the G/G/NBD is relatively large ($s = 90.807$): the customers are quite homogeneous with respect to their mean lifetime. Because of a lack of computational accuracy with the G/G/NBD for two individuals out of 10,330, we used the Gompertz/NBD with parameters $b = 0.750$, $s/\beta = 0.0446$, $r = 1.205$, and $\alpha = 4.430$.

the former model, the expected residual lifetime for each active donor at the end of the first 4.69 years is zero whereas the latter model predicts 7.33 years. A retention policy appears appropriate according to the G/G/NBD whereas the predictions given by the Pareto/NBD suggest focus on the acquisition of new donors.

Simulation

We consider two possible scenarios for the customer lifetime distribution:

(i) *Case of an interior mode.* For a cohort of 3,000 customers, we simulate a G/G/NBD model with the following parameters: $r = 0.75$, $\alpha = 2.5$, $b = 0.35$, $s = 1$, and $\beta = 60$. (The G/G reduces to an exponential/Gompertz model when s is equal to one.) The mode (Equation (17)) equals 11.7 unit periods. For each customer (a) we make a draw u in a uniform distribution over $[0, 1]$ and compute the corresponding lifetime τ :

$$\tau = (1/b) \ln[1 - \beta + \beta/(1 - u)^{1/s}]; \quad (25)$$

(b) we make a draw λ in a gamma distribution with parameters r and α ; (c) we make n repeated draws d_1, d_2, \dots, d_n from an exponential distribution with parameter λ (mean $= 1/\lambda$), where d_j denotes the duration between the $(j - 1)$ th and the j th ($j = 1, 2, \dots, n$) repeat transaction. The number of draws n is such as $d_1 + d_2 + \dots + d_n \leq \tau$; (d) we compute the total number of repeat transactions x_i and the time to the x_i -th repeat transaction t_{xi} during a predetermined period $(0, T]$ with the assumption that all the customers made their initial transaction on the first day of the observation window. We vary T ($T = 5, 10, 15$, and 20 unit periods) and derive the corresponding four vectors (x_i, t_{xi}) for customer i . For each observation period, we estimate the Pareto/NBD model and the G/G/NBD model. The results show that, for all the observation periods, the Pareto distribution reduces to an exponential distribution. The G/G reduces to the G distribution when the length of the observation period is short ($T = 5$). As the observation period becomes long, the three parameters of the G/G distribution are identified. This set of results is compatible with those obtained with the U.S. charity data and the reproduction gift catalog company data in Table 2.

(ii) *Case of a mode at zero.* The parameters of the G/G/NBD model are such as $r = 0.29$, $\alpha = 5.708$, $b = 0.209$, $s = 0.155$, and $\beta = 0.0420$. We partition the observation period into four time intervals of length 10, 20, 40, and 60 time units. Over a broad range of observation periods ($T = 10$ to 20 unit periods), the Pareto/NBD provides a better fit than the G/G/NBD but its relative performance worsens slowly as the

length of the observation period becomes long ($T = 40$ or 60 unit periods).

Overall, we have tested the G/G/NBD versus the Pareto/NBD in six data sets. In the four markets where the mode is away from zero, the G/G/NBD provides a better fit than the Pareto/NBD. (In a fifth market, it improves over the Pareto/NBD when the length of the estimation period is rather long.) At least equally important, the G/G/NBD model can give way to substantially different implications from the Pareto/NBD although the fit and forecasts can be similar. The results seem to suggest that the G/G/NBD tends to be appropriate when firms are reputed and their offerings are differentiated. On the average, the G/G/NBD exhibits slightly more accurate forecasts of the mean number of transactions than the Pareto/NBD. A simulation reproduces the empirical findings.

Conclusion

The study proposes a new model of customer lifetime: the G/G distribution. This model is more flexible than the Pareto distribution. When combined with the NBD, it gives encouraging results with potential differences in the substantive implications from the Pareto/NBD.

Acknowledgments

The authors are grateful to three anonymous reviewers, an associate editor, and Pradeep Chintagunta (the department editor) for their valuable feedback. The authors thank the Direct Marketing Educational Foundation (DMEF) for making data available. Implementation of the model in a spreadsheet and in MATLAB can be found at <http://bit.ly/niABgB>.

Appendix. Likelihood Equations and Probability for a Customer to Be Alive

Individual-Level Likelihood and Probability for a Customer to Be Alive

The individual-level likelihood is such as

$$\begin{aligned} L(\eta, \lambda, b \mid x_i, t_{xi}, T_i) &= \lambda^{x_i} \exp\{-[\eta(e^{bT_i} - 1) + \lambda T_i]\} \\ &\quad + \lambda^{x_i} b \eta e^\eta \int_{t_{xi}}^{T_i} e^{-(\lambda - b)\tau} \exp(-\eta e^{b\tau}) d\tau, \\ &\quad x_i = 0, 1, 2, \dots, 0 \leq t_{xi} < T_i, \\ &= \lambda^{x_i} \exp\{-[\eta(e^{bT_i} - 1) + \lambda T_i]\} \\ &\quad + \lambda^{x_i} \eta e^\eta [e^{t_{xi}(b - \lambda)} E_{\lambda/b}(\eta e^{bt_{xi}}) - e^{T_i(b - \lambda)} E_{\lambda/b}(\eta e^{bT_i})], \quad (26) \end{aligned}$$

where $E_a(z) = \int_1^\infty (e^{-zt}/t^a) dt$ is the exponential integral (Abramowitz and Stegun 1972, p. 228).

When x_i is equal to zero, t_{xi} is equal to zero. The probability for customer i to be alive at time T_i is given by:

$$\begin{aligned} \mathbb{P}(\tau > T_i | \lambda, \eta, b, x_i, t_{xi}, T_i) \\ = L(\lambda | x_i, \tau > T_i) \mathbb{P}(\tau > T_i | \eta, b) / L(\lambda, \eta, b | x_i, t_{xi}, T_i) \\ = \lambda^{x_i} \exp\{-[\eta(e^{bT_i} - 1) + \lambda T_i]\} / L(\lambda, \eta, b | x_i, t_{xi}, T_i). \end{aligned} \quad (27)$$

Aggregate-Level Likelihood and Probability for a Randomly Picked Customer to Be Alive

The aggregate-level likelihood is given by

$$\begin{aligned} L(r, \alpha, b, s, \beta | x_i, t_{xi}, T_i) \\ = (r)_{x_i} [\alpha / (\alpha + T_i)]^r [1 / (\alpha + T_i)]^{x_i} [\beta / (\beta - 1 + e^{bT_i})]^s \\ + (r)_{x_i} b \alpha^r s \beta^s \int_{t_{xi}}^{T_i} (y + \alpha)^{-(r+x_i)} (\beta + e^{by} - 1)^{-(s+1)} e^{by} dy, \end{aligned} \quad (28)$$

where $(r)_{x_i} = \Gamma(r + x_i) / \Gamma(r)$ is the Pochhammer symbol and $\Gamma(\cdot)$ is the gamma function (Abramowitz and Stegun 1972, pp. 255–256). The probability that a randomly picked customer i is alive at time T_i is such as

$$\begin{aligned} \mathbb{P}(\tau > T_i | r, \alpha, b, s, \beta, x_i, t_{xi}, T_i) \\ = L(r, \alpha | x_i, \tau > T_i) \mathbb{P}(\tau > T_i | b, s, \beta) / L(r, \alpha, b, s, \beta | x_i, t_{xi}, T_i) \\ = (r)_{x_i} [\alpha / (\alpha + T_i)]^r [1 / (\alpha + T_i)]^{x_i} [\beta / (\beta - 1 + e^{bT_i})]^s \\ / L(r, \alpha, b, s, \beta | x_i, t_{xi}, T_i). \end{aligned} \quad (29)$$

References

- Abe, M. 2009. "Counting your customers" one by one: A hierarchical Bayes extension to the Pareto/NBD model. *Marketing Sci.* 28(3) 541–553.
- Abramowitz, M., I. A. Stegun. 1972. *Handbook of Mathematical Functions*. Dover Publications, New York.
- Borle, S., S. S. Singh, D. C. Jain. 2008. Customer lifetime value measurement. *Management Sci.* 54(1) 100–112.
- Fader, P. S., B. G. S. Hardie, K. L. Lee. 2005. "Counting your customers" the easy way: An alternative to the Pareto/NBD model. *Marketing Sci.* 24(2) 275–284.
- Fader, P. S., B. G. S. Hardie, J. Shang. 2010. Customer-base analysis in a discrete-time noncontractual setting. *Marketing Sci.* 29(6) 1086–1108.
- Gompertz, B. 1825. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philos. Trans. Roy. Soc. London* 115 513–583.
- Ho, T.-H., Y.-H. Park, Y.-P. Zhou. 2006. Incorporating satisfaction into customer value analysis: Optimal investment in lifetime value. *Marketing Sci.* 25(3) 260–277.
- Jamal, Z., R. E. Bucklin. 2006. Improving the diagnosis and prediction of customer churn: A heterogeneous hazard modeling approach. *J. Interactive Marketing* 20(3/4) 16–29.
- Johnson, N. L., S. Kotz, N. Balakrishnan. 1995. *Continuous Univariate Distributions*, Vol. 2, 2nd ed. John Wiley & Sons, New York.
- Krafft, M. 2002. *Kundenbindung und Kundenwert*. Physica-Verlag, Heidelberg, Germany.
- Malthouse, E. C. 2009. The results from the lifetime value and customer equity modeling competition. *J. Interactive Marketing* 23(3) 272–275.
- Reinartz, W. J., V. Kumar. 2000. On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing. *J. Marketing* 64(4) 17–35.
- Reinartz, W. J., V. Kumar. 2003. The impact of customer relationship characteristics on profitable lifetime duration. *J. Marketing* 67(1) 77–88.
- Schmittlein, D. C., D. G. Morrison, R. Colombo. 1987. Counting your customers: Who are they and what will they do next? *Management Sci.* 33(1) 1–24.
- Sheikh, A. K., J. K. Boah, M. Younas. 1989. Truncated extreme value model for pipeline reliability. *Reliability Engrg. & System Safety* 25(1) 1–14.
- Singh, S. S., S. Borle, D. C. Jain. 2009. A generalized framework for estimating customer lifetime value when customer lifetimes are not observed. *Quant. Marketing Econom.* 7(2) 181–205.