Customer-Base Analysis in a Discrete-Time Noncontractual Setting

Peter S. Fader
The Wharton School of the University of Pennsylvania, Philadelphia, Pennsylvania 19104,
faderp@wharton.upenn.edu

Bruce G. S. Hardie
London Business School, London NW1 4SA, United Kingdom, bhardie@london.edu

Jen Shang
School of Public and Environmental Affairs, Indiana University, Bloomington, Indiana 47405,
jenshang@indiana.edu

Many businesses track repeat transactions on a discrete-time basis. These include (1) companies for whom transactions can only occur at fixed regular intervals, (2) firms that frequently associate transactions with specific events (e.g., a charity that records whether supporters respond to a particular appeal), and (3) organizations that choose to utilize discrete reporting periods even though the transactions can occur at any time. Furthermore, many of these businesses operate in a noncontractual setting, so they have a difficult time differentiating between those customers who have ended their relationship with the firm versus those who are in the midst of a long hiatus between transactions. We develop a model to predict future purchasing patterns for a customer base that can be described by these structural characteristics. Our beta-geometric/beta-Bernoulli (BG/BB) model captures both of the underlying behavioral processes (i.e., customers' purchasing while "alive" and time until each customer permanently "dies"). The model is easy to implement in a standard spreadsheet environment and yields relatively simple closed-form expressions for the expected number of future transactions conditional on past observed behavior (and other quantities of managerial interest). We apply this discrete-time analog of the well-known Pareto/NBD model to a data set on donations made by the supporters of a nonprofit organization located in the midwestern United States. Our analysis demonstrates the excellent ability of the BG/BB model to describe and predict the future behavior of a customer base.

Key words: BG/BB; beta-geometric; beta-binomial; customer-base analysis; customer lifetime value; CLV; RFM; Pareto/NBD
History: Received: March 24, 2009; accepted: March 31, 2010; accepted by Scott A. Neslin, acting editor-in-chief. Published online in *Articles in Advance* August 11, 2010.

1. Introduction

Consider a major nonprofit organization located in the midwestern United States that is funded in large part by donations from individuals. In 1995 the organization "acquired" 11,104 first-time supporters; in each of the following six years, these individuals either did or did not support the organization. As shown in Table 1, donation behavior can be characterized by a binary string, where 1 indicates that a donation was made. (For the purposes of this analysis—similar to Netzer et al. 2008—we focus only on the annual incidence on the donations; we ignore the dollar values.) Given these data, management would like to know which individuals are most likely to be active donors in the future so that it can predict the level of "transactions" it can expect in future years from this cohort of donors (both individually and collectively).

Management has a five-year planning period and therefore would like to forecast the expected number of donations for the 1995 cohort as a whole, as well as for particular types of individuals, over the period 2002–2006. For instance,

• What should be expected from donor 100008, who has made a repeat donation in each of the six years since becoming a supporter of the organization: is he likely to go "five-for-five" in the future period? If not, how much "shrinkage" would we expect?

• How about comparing donor 100009, who had been a consistent supporter up until 2001, versus donor 100004, who has had a more irregular history, with one fewer donation overall but with one made in 2001?

• Likewise, how does donor 100004 compare to donor 111103? They have both made four repeat donations, including one in 2001, but their earlier histories differ somewhat from each other.

• Finally, how about the many donors (such as 100001) who have done nothing since their initial contributions? Should the nonprofit organization write

Table 1   Annual Donation Behavior by the 1995 Cohort of First-Time Supporters

ID	1995	1996	1997	1998	1999	2000	2001
100001	1	0	0	0	0	0	0
100002	1	0	0	0	0	0	0
100003	1	0	0	0	0	0	0
100004	1	0	1	0	1	1	1
100005	1	0	1	1	1	0	1
100006	1	1	1	1	0	1	0
100007	1	1	0	1	0	1	0
100008	1	1	1	1	1	1	1
100009	1	1	1	1	1	1	0
100010	1	0	0	0	0	0	0
⋮		⋮		⋮			
111102	1	1	1	1	1	1	1
111103	1	0	1	1	0	1	1
111104	1	0	0	0	0	0	0

them off, or is there still some meaningful future value in them—individually and collectively?

Recognizing that this a noncontractual setting,[1] the marketing analyst may think, "Let's use the Pareto/NBD," a model developed by Schmittlein et al. (1987) to provide answers to the kinds of customer-base analysis questions listed above.

But is this an appropriate way to proceed? At the heart of the Pareto/NBD model is the assumption that customer purchasing while "alive" is characterized by a Poisson distribution and that cross-sectional heterogeneity in the mean purchase rates is characterized by a gamma distribution (resulting in the negative binomial distribution (NBD) model of repeat buying; Ehrenberg 1988, Morrison and Schmittlein 1988). The use of the Poisson distribution assumes that transactions can occur at any point in time; this may be an acceptable assumption for the purchasing of CDs from a website or for the purchasing of office products in a business-to-business (B2B) setting, which are the empirical settings considered by Fader et al. (2005) and Schmittlein and Peterson (1994), respectively. However, it is not a valid assumption in a number of other situations, including the nonprofit setting described above. Even Schmittlein et al. (1987) acknowledge that their model has limited applicability and that there is a need for an alternative modeling

[1] In a contractual setting (e.g., gym membership, cable TV, theater subscription plan), we observe the time at which the customer "dies" (i.e., ends their formal relationship with the firm). In a non-contractual setting (e.g., traditional mail order, retail store patronage), however, the time at which a customer dies is unobserved by the firm; customers do not notify the firm "when they stop being a customer. Instead they just silently attrite" (Mason 2003, p. 55). The only potential evidence of this having happened is an unusually long hiatus since the last recorded purchase. The challenge facing the analyst is how to differentiate between those customers who have ended their relationship with the firm versus those who are simply in the midst of a long hiatus between transactions.

framework to accommodate business settings characterized by discrete-time purchasing (see pp. 16–17 and Table 3 in their paper), yet no one to date has presented such a model.

As another example, consider attendance at the INFORMS Marketing Science Conference. The conference occurs at a discrete point in time and an individual can either attend or not. Similarly, consider Sunday church attendance; an individual can either attend the Sunday morning service or not. In both cases, the opportunities for a transaction occur at discrete points in time, and there is an upper bound on the number of transactions that can occur in a fixed unit of time; an individual cannot attend the INFORMS Marketing Science Conference more than once a year or attend the Sunday morning church service more than 52 times a year. In such noncontractual settings, the behavior is "necessarily discrete," and it is clearly incorrect to model the number of transactions using a Poisson distribution. It would be more appropriate to model the number of transactions in a given time period using a Bernoulli process.

In other settings, the behavior of interest can occur in continuous time, but it is "effectively discrete" in the way firms view it. Consider the case of blood donations. A blood collection agency will send quarterly notices to its donor base, requesting that they give blood. Although an individual can give blood at any point in time during that quarter, there is still an upper bound in the number of times the agency is willing to accept blood from any donor and can therefore characterize a donor's behavior in terms of whether or not she gave blood in a fixed time interval. Similarly, a charity may send out letters every six months requesting money. Although an individual can send in a donation at any point in time, the charity is basically interested in whether or not he responded to a specific request for funds and will therefore characterize donation behavior simply in terms of whether or not the individual responds to a mailing (Piersma and Jonker 2004). A number of mail-order companies also think of their customer behavior in such a manner (e.g., did the customer place an order in response to the quarterly catalog mailing?). In these cases, it is convenient to think of there being a natural upper bound on the number of transactions that can occur in a fixed unit of time (e.g., year), and it is therefore more appropriate to model the number of transactions using a Bernoulli process rather than a Poisson distribution.

Finally, there are cases where the event of interest has no constraints on it at all—it is truly a continuous-time behavior, but it is so rare per unit of time that management will choose to discretize the purchasing data for analysis and reporting purposes. For example, a cruise-ship company may characterize customer

behavior in terms of whether or not each customer went on a cruise in 2000, 2001, 2002, etc. (Berger et al. 2003). Once again, purchasing behavior is more conveniently described as a Bernoulli process rather than as a Poisson process. An example of this in a consumer packaged goods setting is the work of Chatfield and Goodhardt (1970), who model the purchasing of a product not in terms of the number of purchases made by an individual in a 24-week period (using the NBD model) but rather in terms of the number of weeks in which an individual purchased the product (using the beta-binomial model of Skellam 1948, with $n = 24$). Similarly, Easton (1980) uses the beta-binomial model to characterize purchasing in an industrial setting, commenting that using a discrete purchase interval is a useful way of overcoming the problem of determining when exactly a purchase is deemed to have occurred in a B2B setting.

Figure 1 illustrates this continuum of settings in which it is either correct or simply makes more sense to model individual-level transaction behavior using a Bernoulli process rather than a Poisson distribution. In all of these settings, it is inappropriate to use the Pareto/NBD as the underlying model for a customer-base analysis exercise.

In this paper we develop a model that can be used to answer the critical customer-base analysis questions in discrete-time, noncontractual settings; in other words, we develop a discrete-time analog of the Pareto/NBD model. Although many aspects of the Pareto/NBD model (and the inferences frequently associated with it) carry over fairly smoothly to the discrete-time setting, there are a number of interesting issues that arise in the discrete-time setting that are quite unique—and offer significant benefits for model implementation. In the next section, we first outline the assumptions underpinning this model and then present expressions for a number of managerially relevant quantities. This is followed by an empirical analysis (for the aforementioned nonprofit organization) in which we carefully examine the performance of the model both in a six-year calibration sample and a five-year

holdout period. We then examine the relative performance of the Pareto/NBD model when applied to this same data set. Next we present an extension to the basic model in which the consequences of relaxing one of the model assumptions are explored. We conclude with a discussion of several additional issues that arise from this work.

2. Model Development

Our objective is to develop a stochastic model of buyer behavior for discrete-time, noncontractual settings. To start, we define a *transaction opportunity* as either one of the following:

• A well-defined *point in time* at which a transaction either occurs or does not occur, or

• A well-defined *time interval* during which a transaction either occurs or does not occur.

The first type of transaction opportunity corresponds to the necessarily discrete case in Figure 1. The second type of transaction opportunity corresponds to the "generally discrete" and "discretized by recording process" cases in Figure 1. (The nonprofit example discussed in the introduction is an example of this second case.) In all three cases, a customer's transaction history can be expressed as a binary string, where $y_t = 1$ if a transaction occurred at or during the tth transaction opportunity, and 0 otherwise (for $t = 1, \ldots, n$ transaction opportunities). Note that we are simply interested in modeling the transaction process (i.e., the pattern of 1s and 0s). We are not interested in modeling other behaviors associated with each transaction (e.g., the quantity purchased); this is discussed in §6.

Our model is based on the following six assumptions.

ASSUMPTION 1. *A customer's relationship with the firm has two phases: he is "alive" (A) for some period of time, then becomes permanently inactive ("dies"; D).*

ASSUMPTION 2. *While alive, the customer buys at any given transaction opportunity with probability p:*

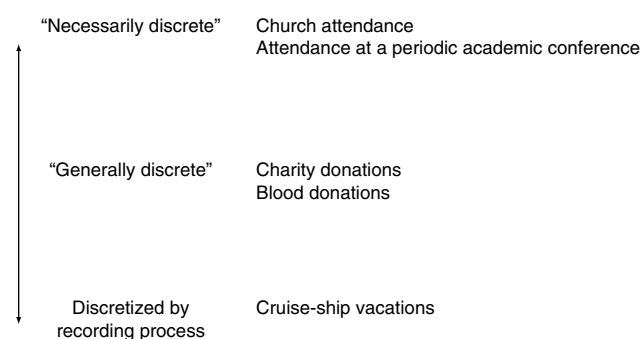$$P(Y_t = 1 \mid p, \text{alive at } t) = p, \quad 0 \le p \le 1.$$

(*This implies that the number of transactions by a customer alive for i transaction opportunities follows a binomial (i, p) distribution.*)

ASSUMPTION 3. *A "living" customer dies at the beginning of a transaction opportunity with probability θ. (This implies that the (unobserved) lifetime of a customer is characterized by a geometric distribution.)*

ASSUMPTION 4. *Heterogeneity in p follows a beta distribution with probability distribution function (pdf)*

$$f(p \mid \alpha, \beta) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 \le p \le 1, \ \alpha, \beta > 0. \quad (1)$$

Figure 1 Classifying "Discrete-Time" Transaction Opportunities

"Necessarily discrete"	Church attendance
	Attendance at a periodic academic conference
"Generally discrete"	Charity donations
	Blood donations
Discretized by recording process	Cruise-ship vacations

ASSUMPTION 5. *Heterogeneity in θ follows a beta distribution with pdf*

$$f(\theta \mid \gamma, \delta) = \frac{\theta^{\gamma-1}(1-\theta)^{\delta-1}}{B(\gamma, \delta)}, \quad 0 \le \theta \le 1, \ \gamma, \delta > 0. \quad (2)$$

ASSUMPTION 6. *The transaction probability p and the dropout probability θ vary independently across customers.*

Assumptions (2) and (4) yield the beta-Bernoulli model (i.e., the beta-binomial model without the binomial coefficient, since we explicitly account for the ordering of the transactions). Similarly, Assumptions (3) and (5) yield the beta-geometric (BG) distribution. We therefore call this the beta-geometric/beta-Bernoulli (BG/BB) model of buyer behavior.

2.1. Derivation of Model Likelihood Function

Consider a customer with repeat purchase string 10100. What is $P(Y_1 = 1, Y_2 = 0, Y_3 = 1, Y_4 = 0, Y_5 = 0 \mid p, \theta)$? The fact that the customer made a purchase at the third transaction opportunity means that he must have been alive for $t = 1, 2, 3$. However, $Y_4 = 0, Y_5 = 0$ could be the result of one of three scenarios: (i) he died at the beginning of the fourth transaction opportunity (AAADD), (ii) he was alive at the fourth transaction opportunity and died at the beginning of the fifth transaction opportunity (AAAAD), or (iii) he was alive at both the fourth and fifth transaction opportunities (AAAAA). We therefore compute $P(Y_1 = 1, Y_2 = 0, Y_3 = 1, Y_4 = 0, Y_5 = 0 \mid p, \theta)$ by computing the probability of the purchase string conditional on each scenario and multiplying it by the probability of that scenario:

$f(10100 \mid p, \theta)$

$\quad = f(10100 \mid p, \text{AAADD})P(\text{AAADD} \mid \theta)$

$\qquad + f(10100 \mid p, \text{AAAAD})P(\text{AAAAD} \mid \theta)$

$\qquad + f(10100 \mid p, \text{AAAAA})P(\text{AAAAA} \mid \theta)$

$\quad = p(1-p)p \underbrace{(1-\theta)^3\theta}_{P(\text{AAADD})} + p(1-p)p(1-p)\underbrace{(1-\theta)^4\theta}_{P(\text{AAAAD})}$

$\qquad + \underbrace{p(1-p)p}_{P(Y_1=1,Y_2=0,Y_3=1)}(1-p)(1-p)\underbrace{(1-\theta)^5}_{P(\text{AAAAA})}. \quad (3)$

Note that the zero-order nature of purchasing while the customer is alive means that the exact order of any given number of transactions prior to the last observed transaction does not matter. For example, it should be clear that $f(10100 \mid p, \theta) = f(01100 \mid p, \theta)$. Therefore, we do not need the complete binary-string representation of a customer's transaction history. Rather, all we need to know for n transaction opportunities are *frequency* and *recency*: the number of transactions across the calibration period ($x = \sum_{t=1}^n y_t$) and the transaction opportunity at which the last

observed transaction occurred (t_x).[2] We therefore go from 2^n binary string representations of all the possible purchase patterns to $n(n + 1)/2 + 1$ possible recency/frequency patterns.

This realization that recency and frequency are sufficient summary statistics offers signficant benefits for model implementation, particularly as the number of transaction opportunities becomes sizeable. For instance, in the case of our nonprofit organization, we can compress the number of necessary binary strings from 64 down to 22 recency/frequency combinations, making it a bit easier to visualize and manipulate the data set. However, in another recent application with $n = 10$, we saw a reduction from 1,024 binary strings down to 56 recency/frequency combinations. Furthermore, these numbers are not affected by the size of the customer base being modeled; see Table 2 for a complete characterization of the nonprofit data set partially presented in Table 1. Whether we have 11,000 customers or 11 million customers, the data structure would be identical—the numbers in the "No. of donors" columns would grow, but the computational demands for data storage and manipulation would be unaffected.

Returning to the likelihood function, we generalize the logic behind the construction of (3), so it follows that

$L(p, \theta \mid x, t_x, n) = p^x(1-p)^{n-x}(1-\theta)^n$

$$+ \sum_{i=0}^{n-t_x-1} p^x(1-p)^{t_x-x+i}\theta(1-\theta)^{t_x+i}. \quad (4)$$

To arrive at the likelihood function for a randomly chosen customer with purchase history (x, t_x, n), we remove the conditioning on p and θ by taking the expectation of (4) over their respective mixing distributions:

$L(\alpha, \beta, \gamma, \delta \mid x, t_x, n)$

$\quad = \int_0^1 \int_0^1 L(p, \theta \mid x, t_x, n) f(p \mid \alpha, \beta) f(\theta \mid \gamma, \delta) \, dp \, d\theta$

$\quad = \frac{B(\alpha+x, \beta+n-x)}{B(\alpha, \beta)} \frac{B(\gamma, \delta+n)}{B(\gamma, \delta)}$

$\qquad + \sum_{i=0}^{n-t_x-1} \frac{B(\alpha+x, \beta+t_x-x+i)}{B(\alpha, \beta)}$

$\qquad \cdot \frac{B(\gamma+1, \delta+t_x+i)}{B(\gamma, \delta)}. \quad (5)$

(The solution to the double integral follows naturally from the integral representation of the beta function.)

[2] If $x = 0$, then $t_x = 0$. Note that this measure of recency differs from that normally used by the direct marketing community, who measure recency as the time from the last observed transaction to the end of the observation period (i.e., $n - t_x$).

Table 2   Recency/Frequency Summary of the Annual Donation Behavior by the 1995 Cohort of First-Time Supporters ($n = 6$)

x	t_x	No. of donors	x	t_x	No. of donors
6	6	1,203	4	4	240
5	6	728	3	4	181
4	6	512	2	4	155
3	6	357	1	4	78
2	6	234	3	3	322
1	6	129	2	3	255
5	5	335	1	3	129
4	5	284	2	2	613
3	5	225	1	2	277
2	5	173	1	1	1,091
1	5	119	0	0	3,464

The four BG/BB model parameters $(\alpha, \beta, \gamma, \delta)$ can be estimated via the method of maximum likelihood in the following manner. For a calibration period with n transaction opportunities, we have $J = n(n+1)/2+1$ possible recency/frequency patterns, each containing f_j customers. The sample log-likelihood function is given by

$$LL(\alpha, \beta, \gamma, \delta) = \sum_{j=1}^{J} f_j \ln[L(\alpha, \beta, \gamma, \delta \mid x_j, t_{x_j}, n)], \quad (6)$$

where x_j and t_{x_j} are the frequency and recency, respectively, for each unique pattern. This can be maximized using standard numerical optimization routines. These calculations are easy to perform in a spreadsheet environment; in fact, the entire model implementation (from initial data setup through the calculation of the "key results" in the next section) rarely requires the analyst to use any software beyond a spreadsheet. This is a major benefit of the BG/BB model.

2.2. Key Results

We now present expressions for a set of quantities of interest to anyone wanting to apply this model of buyer behavior in a discrete-time, noncontractual setting. (The associated derivations can be found in Appendix A.)

Let the random variable $X(n) = \sum_{t=1}^{n} Y_t$ denote the number of transactions occurring across the first n transaction opportunities. The BG/BB probability mass function is

$$P(X(n) = x \mid \alpha, \beta, \gamma, \delta)$$
$$= \binom{n}{x} \frac{B(\alpha+x, \beta+n-x)}{B(\alpha, \beta)} \frac{B(\gamma, \delta+n)}{B(\gamma, \delta)}$$
$$+ \sum_{i=x}^{n-1} \binom{i}{x} \frac{B(\alpha+x, \beta+i-x)}{B(\alpha, \beta)} \frac{B(\gamma+1, \delta+i)}{B(\gamma, \delta)}, \quad (7)$$

with mean

$$E(X(n) \mid \alpha, \beta, \gamma, \delta)$$
$$= \left(\frac{\alpha}{\alpha+\beta}\right)\left(\frac{\delta}{\gamma-1}\right)$$
$$\cdot \left\{1 - \frac{\Gamma(\gamma+\delta)}{\Gamma(\gamma+\delta+n)} \frac{\Gamma(1+\delta+n)}{\Gamma(1+\delta)}\right\}. \quad (8)$$

More generally, let the random variable $X(n, n+n^*) = \sum_{t=n+1}^{n^*} Y_t$ denote the number of transactions in the interval $(n, n + n^*]$. The BG/BB probability of x^* transactions occurring in this interval is given by

$$P(X(n, n+n^*) = x^* \mid \alpha, \beta, \gamma, \delta)$$
$$= \delta_{x^*=0} \left\{1 - \frac{B(\gamma, \delta+n)}{B(\gamma, \delta)}\right\}$$
$$+ \binom{n^*}{x^*} \frac{B(\alpha+x^*, \beta+n^*-x^*)}{B(\alpha, \beta)} \frac{B(\gamma, \delta+n+n^*)}{B(\gamma, \delta)}$$
$$+ \sum_{i=x^*}^{n^*-1} \binom{i}{x^*} \frac{B(\alpha+x^*, \beta+i-x^*)}{B(\alpha, \beta)} \frac{B(\gamma+1, \delta+n+i)}{B(\gamma, \delta)}, \quad (9)$$

with mean

$$E(X(n, n+n^*) \mid \alpha, \beta, \gamma, \delta)$$
$$= \left(\frac{\alpha}{\alpha+\beta}\right)\left(\frac{\delta}{\gamma-1}\right)\frac{\Gamma(\gamma+\delta)}{\Gamma(1+\delta)}$$
$$\times \left\{\frac{\Gamma(1+\delta+n)}{\Gamma(\gamma+\delta+n)} - \frac{\Gamma(1+\delta+n+n^*)}{\Gamma(\gamma+\delta+n+n^*)}\right\}. \quad (10)$$

In most customer-base analysis settings, we are interested in making statements about customers conditional on their observed purchase history (x, t_x, n).

• The probability that a customer with purchase history (x, t_x, n) will be alive at the $(n+1)$th transaction opportunity is

$$P(\text{alive at } n+1 \mid \alpha, \beta, \gamma, \delta, x, t_x, n)$$
$$= \frac{B(\alpha+x, \beta+n-x)}{B(\alpha, \beta)} \frac{B(\gamma, \delta+n+1)}{B(\gamma, \delta)}$$
$$\cdot L(\alpha, \beta, \gamma, \delta \mid x, t_x, n)^{-1}. \quad (11)$$

• The probability that a customer with purchase history (x, t_x, n) makes x^* transactions in the interval $(n, n+n^*]$ is

$$P(X(n, n+n^*) = x^* \mid \alpha, \beta, \gamma, \delta, x, t_x, n)$$
$$= \delta_{x^*=0} \left\{1 - \frac{\mathscr{A}_1}{L(\alpha, \beta, \gamma, \delta \mid x, t_x, n)}\right\}$$
$$+ \frac{\mathscr{A}_2}{L(\alpha, \beta, \gamma, \delta \mid x, t_x, n)}, \quad (12)$$

where

$$\mathscr{A}_1 = \frac{B(\alpha+x, \beta+n-x)}{B(\alpha, \beta)} \frac{B(\gamma, \delta+n)}{B(\gamma, \delta)}$$

and

$$\mathscr{A}_2 = \binom{n^*}{x^*} \frac{B(\alpha+x+x^*, \beta+n-x+n^*-x^*)}{B(\alpha, \beta)}$$
$$\cdot \frac{B(\gamma, \delta+n+n^*)}{B(\gamma, \delta)}$$
$$+ \sum_{i=x^*}^{n^*-1} \binom{i}{x^*} \frac{B(\alpha+x+x^*, \beta+n-x+i-x^*)}{B(\alpha, \beta)}$$
$$\cdot \frac{B(\gamma+1, \delta+n+i)}{B(\gamma, \delta)}.$$

- The expected number of future transactions across the next n^* transaction opportunities by a customer with purchase history (x, t_x, n) is

$$E(X(n, n+n^*) \mid \alpha, \beta, \gamma, \delta, x, t_x, n)$$
$$= \frac{1}{L(\alpha, \beta, \gamma, \delta \mid x, t_x, n)} \frac{B(\alpha+x+1, \beta+n-x)}{B(\alpha, \beta)}$$
$$\times \left(\frac{\delta}{\gamma-1}\right) \frac{\Gamma(\gamma+\delta)}{\Gamma(1+\delta)}$$
$$\cdot \left\{ \frac{\Gamma(1+\delta+n)}{\Gamma(\gamma+\delta+n)} - \frac{\Gamma(1+\delta+n+n^*)}{\Gamma(\gamma+\delta+n+n^*)} \right\}. \quad (13)$$

Many customer-base analysis exercises are motivated by a desire to compute customer lifetime value (CLV), which is "the present value of the future cash flows attributed to the customer relationship" (Pfeifer et al. 2005, p. 17). The general explicit formula for computing CLV is (Rosset et al. 2003)

$$E(\text{CLV}) = \int_0^\infty E[v(t)] S(t) d(t) \, dt,$$

where $E[v(t)]$ is the expected value of the customer at time t (assuming he is alive), $S(t)$ is the survivor function, and $d(t)$ is a discount factor that reflects the present value of money received at time t. Following Fader et al. (2005), if we assume that the process describing the net cash flow per transaction for a given customer is both independent of the transaction process and stationary, we can express $v(t)$ as net cash flow/transaction $\times t(t)$, where $t(t)$ is the transaction rate at t.

In many cases we are interested in the expected *residual* lifetime value of a customer. Standing at time T,

$$E(\text{RLV}) = E(\text{net cashflow/transaction})$$
$$\times \underbrace{\int_T^\infty E[t(t)]S(t \mid t > T)d(t - T) \, dt}_{\text{discounted expected residual transactions}}.$$

The number of discounted expected residual transactions (DERT) is the present value of the expected future transaction stream for a customer with purchase history (x, t_x, T). Fader et al. (2005) derive the expression for this quantity when the transaction process can be described by the Pareto/NBD model. When the transaction process is described by the BG/BB model, the present value of the expected number of future transactions for a customer with purchase history (x, t_x, n), with discount rate d is

$$\text{DERT}(d \mid \alpha, \beta, \gamma, \delta, x, t_x, n)$$
$$= \frac{B(\alpha+x+1, \beta+n-x)}{B(\alpha, \beta)} \frac{B(\gamma, \delta+n+1)}{B(\gamma, \delta)(1+d)}$$
$$\times \frac{{}_2F_1(1, \delta+n+1; \gamma+\delta+n+1; 1/(1+d))}{L(\alpha, \beta, \gamma, \delta \mid x, t_x, n)}, \quad (14)$$

where ${}_2F_1(\cdot)$ is the Gaussian hypergeometric function.[3] This number of discounted expected residual transactions can then be rescaled by the customer's value multiplier to yield an overall estimate of $E(\text{RLV})$. Although the presence of the Gaussian hypergeometric function makes this calculation a bit more complex than the others in this section, it is worth emphasizing that it only needs to be evaluated once for any given value of n (i.e., only once per cohort, not for every recency/frequency pattern), and it is relatively straightforward to use a recursion formula to perform the calculations in a spreadsheet environment. Furthermore, this calculation for DERT is far simpler than the equivalent expression derived by Fader et al. (2005) for the Pareto/NBD model. In that case, the DERT expression required the evaluation of Gaussian hypergeometric functions for each recency/frequency combination, as well as the confluent hypergeometric function of the second kind, which is unfamiliar and fairly burdensome from a computational standpoint.

Finally, we may also be interested in making inferences about a customer's latent transaction and dropout probabilities.

- The marginal posterior distribution of P is

$$f(p \mid \alpha, \beta, \gamma, \delta, x, t_x, n) = \mathscr{B} / L(\alpha, \beta, \gamma, \delta \mid x, t_x, n), \quad (15)$$

where

$$\mathscr{B} = \frac{p^{\alpha+x-1}(1-p)^{\beta+n-x-1}}{B(\alpha, \beta)} \frac{B(\gamma, \delta+n)}{B(\gamma, \delta)}$$
$$+ \sum_{i=0}^{n-t_x-1} \frac{p^{\alpha+x-1}(1-p)^{\beta+t_x-x+i-1}}{B(\alpha, \beta)} \frac{B(\gamma+1, \delta+t_x+i)}{B(\gamma, \delta)}.$$

[3] Assuming that there are k transaction opportunities per year, an annual discount rate of r maps to a discount rate of $d = (1+r)^{1/k} - 1$.

- The marginal posterior distribution of Θ is

$$f(\theta \mid \alpha, \beta, \gamma, \delta, x, t_x, n) = \mathscr{C}/L(\alpha, \beta, \gamma, \delta \mid x, t_x, n), \quad (16)$$

where

$$\mathscr{C} = \frac{B(\alpha + x, \beta + n - x)}{B(\alpha, \beta)} \frac{\theta^{\gamma - 1}(1 - \theta)^{\delta + n - 1}}{B(\gamma, \delta)}$$

$$+ \sum_{i=0}^{n - t_x - 1} \frac{B(\alpha + x, \beta + t_x - x + i)}{B(\alpha, \beta)} \frac{\theta^{\gamma}(1 - \theta)^{\delta + t_x + i - 1}}{B(\gamma, \delta)}.$$

- For $l, m = 0, 1, 2, \ldots$, the (l, m)th product moment of the joint posterior distribution of P and Θ is

$$E(P^l \Theta^m \mid \alpha, \beta, \gamma, \delta, x, t_x, n)$$

$$= \frac{B(\alpha + l, \beta)}{B(\alpha, \beta)} \frac{B(\gamma + m, \delta)}{B(\gamma, \delta)}$$

$$\cdot \frac{L(\alpha + l, \beta, \gamma + m, \delta \mid x, t_x, n)}{L(\alpha, \beta, \gamma, \delta \mid x, t_x, n)}, \quad (17)$$

where $L(\alpha + l, \beta, \gamma + m, \delta \mid x, t_x, n)$ is simply (5) evaluated using $\alpha + l$ in place of α and $\gamma + m$ in place of γ.

3. Empirical Analysis

We examine the performance of the BG/BB model using data on the annual donation behavior by the supporters of a nonprofit organization located in the midwestern United States. The full data set contains information on the 56,847 people who made their first-ever annual donation between 1995 and 2000 (inclusive), from their first year up to and including 2006; the sizes of each annual cohort are given in Table 3.

Our initial analysis focuses on the 11,104 members of the 1995 cohort. We fit the model using the data on whether or not these supporters made repeat donations across 1996–2001 and examine the model's predictive performance across a 2002–2006 holdout validation period. We follow up this analysis with one in which we pool the six cohorts, fitting the model to the repeat donation data up to and including 2001 and examining its predictive performance over 2002–2006. (For the sake of linguistic simplicity, we will refer to the act of making a repeat donation in any given year as making a repeat transaction or purchase.)

3.1. Analysis of the 1995 Cohort

The group of 11,104 people that became supporters of the organization for the first time in 1995 made a total of 24,615 repeat transactions over the next six years. Given the data in Figure 2, we "code up" the log-likelihood function given in (6) in Excel—see Figure 2 for a screensheet of the complete spreadsheet used for parameter estimation—and maximize it using the Solver add-in. (A note on how to implement the model in Excel, along with a copy of the complete spreadsheet, can be found at http://brucehardie.com/notes/010/.) The resulting maximum-likelihood estimates of the model parameters are reported in Table 4. (We also report the model parameters and value of the log-likelihood function for the beta-Bernoulli model and note that the addition of the "death" component results in a major improvement in model fit.)

The expected number of people making $0, 1, \ldots, 6$ repeat transactions between 1996 and 2001 is computed using (7) and compared to the actual frequency distribution in Figure 3. We note that the model provides a very good fit to the data.

The performance of the model becomes more impressive when we see how well it tracks repeat transactions over time. Using the expression for the expected number of transactions across n transaction opportunities as given in (8), we compute the expected number of repeat transactions made by the whole cohort of 11,104 people up to 2006. These are plotted along with the actual cumulative numbers in Figure 4(a). We note that the BG/BB model predictions accurately track the actual cumulative number of repeat transactions in both the six-year calibration period and the five-year forecast period, underforecasting at 2006 by a mere -0.65%.[4] Further insight into the excellent tracking performance of the model is given in Figure 4(b), which reports these numbers on a year-by-year basis; we note that the BG/BB model clearly captures the underlying trend in repeat transactions over this fairly lengthy period of time.

To get a clearer idea of how well the model captures validation period purchasing, we compute the expected number of people making $x^* = 0, 1, \ldots, 5$ transactions in 2002–2006 ($n^* = 5$) using (9) and compare it to the actual frequency distribution in Figure 5. We note that the model provides a very good prediction of the actual behavior.

3.1.1. Conditional Expectations. Perhaps a more important examination of the predictive performance of the model focuses on the quality of the predictions of future behavior conditional on past behavior.

Table 3	Number of New Supporters Each Year (1995–2000)	
Cohort	Size
1995	11,104
1996	10,057
1997	9,043
1998	8,175
1999	8,977
2000	9,491

---

[4] As a point of comparison, the prediction associated with the BB model overforecasts cumulative repeat transactions at the end of 2006 by 20%.

Figure 2   Screenshot of Excel Worksheet Used for Parameter Estimation

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	alpha	1.204	B(alpha, beta)		1.146									
2	beta	0.750												
3	gamma	0.657	B(gamma, delta)		0.729									
4	delta	2.783												
5														
6	LL	−33,225.6												
7														
8	x	t_x	n	# donors	L(.|x = x, t_x, n)		n -t_x − 1		0	1	2	3	4	5
9	6	6	6	1,203	−2,624.6	0.1129	−1	0.1129	0	0	0	0	0	0
10	5	6	6	728	−3,126.7	0.0136	−1	0.0136	0	0	0	0	0	0
11	4	6	6	512			−1		0	0	0	0	0	0
12	3	6	6	357			−1		0	0	0	0	0	0
13	2	6	6	234	−1,322.5	0.0035	−1	0.0035	0	0	0	0	0	0
14	1	6	6	129	−630.0	0.0076	−1	0.0076	0	0	0	0	0	0
15	5	5	6	335	−1,245		0	0.0036	0.0107	0	0	0	0	0
16	4	5	6	284	−1,447.1	0.0061	0	0.0046	0.0015	0	0	0	0	0
17	3	5	6		−663.5	0.0036	0	0.0030	0.0006	0	0	0	0	0
18	2	5	6	173	−952.6	0.0041	0	0.0035	0.0005	0	0	0	0	0
19	1	5	6	119	−567.3	0.0085	0		0	0	0	0	0	0
20	4	4	6	240	−923.6	0.0213	1	0.0046	0.0152	0.0015	0	0	0	0
21	3	4	6	181	−915.7	0.0063	1	0.0030	0.0027	0.0006	0	0	0	0
22	2	4	6	155	−805.3	0.0055	1	0.0035	0.0015	0.0005	0	0	0	0
23	1	4	6	78	−356.5	0.0104	1	0.0076	0.0018	0.0009	0	0	0	0
24	3	3	6	322	−1,135.8	0.0294	2	0.0030	0.0230	0.0027	0.0006	0	0	0
25	2	3	6	255	−1,151.6	0.0109	2	0.0035	0.0054	0.0015	0.0005	0	0	0
26	1	3	6	129	−545.0	0.0146	2	0.0076	0.0043	0.0018	0.0009	0	0	0
27	2	2	6	613	−1,846.4	0.0492	3	0.0035	0.0383	0.0054	0.0015	0.0005	0	0
28	1	2	6	277	−993.9	0.0276	3	0.0076	0.0130	0.0043	0.0018	0.0009	0	0
29	1	1	6	1,091	−2,497.1	0.1014	4	0.0076	0.0737	0.0130	0.0043	0.0018	0.0009	0
30	0	0	6	3,464	−4,044.3	0.3111	5	0.0362	0.1909	0.0459	0.0189	0.0098	0.0058	0.0037

Formula callouts:
- E1: `=EXP(GAMMALN(B1)+GAMMALN(B2)−GAMMALN(B1+B2))`
- (E3/E4): `=EXP(GAMMALN($B$1+A9)+GAMMALN($B$2+C9−A9)−GAMMALN($B$1+$B$2+C9))/$E$1*EXP(GAMMALN($B$3)+GAMMALN($B$4+C9)−GAMMALN($B$3+$B$4+C9))/$E$3`
- B6: `=SUM(E9: E30)`
- (H9): `=IF(I$8<=$G9,EXP(GAMMALN($B$1+A9)+GAMMALN($B$2+$B9−$A9+I$8)−GAMMALN($B$1+$B$2+$B9+I$8))/$E$1*EXP(GAMMALN($B$3+1)+GAMMALN($B$4+$B9+I$8)−GAMMALN($B$3+$B$4+$B9+I$8+1))/$E$3,0)`
- G15: `=C15-B15-1`
- E17: `=D19*LN(F19)`
- F19: `=SUM(H19: N19)`

We use (13) to compute the expected number of transactions in the 2002–2006 period ($n^* = 5$) conditional on each of the 22 (x, t_x) patterns associated with $n = 6$. These conditional expectations are reported in Table 5 as a function of recency (the year of the individual's last transaction) and frequency (the number of repeat transactions).

In Figure 6(a) we report these conditional expectations, along with the average of the number of the transactions that actually occurred in the 2002–2006 forecast period, broken down by the number of repeat transactions in 1996–2001. (For each x, we are averaging over customers with different values of t_x.) Similarly, Figure 6(b) reports these conditional expectations along with the average of the number of the transactions that actually occurred in the 2002–2006 forecast period, broken down by the year of the individual's last transaction. (For each t_x, we are averaging over customers with different values of x.) We observe that the BG/BB model generates very good predictions of the expected behavior in the longitudinal holdout period, with the only real blemish

being an underestimation of expected purchasing by those individuals whose last repeat purchase occurred before 1998.

Referring back to Table 5, we can now address the questions about different kinds of customers raised at the outset of the paper.

• A donor who has made a repeat transaction every year is expected to make "only" 3.75 transactions over the next five years. Of course, such donors are still extremely valuable, but the possibility of death plus the fact that they might have been somewhat lucky in the past make them a bit less valuable than they might

Table 4   Parameter Estimates, 1995 Cohort

	α	β	γ	δ	LL
BB	0.487	0.826			−35,516.1
BG/BB	1.204	0.750	0.657	2.783	−33,225.6

Figure 3   Predicted vs. Actual Frequency of Repeat Transactions

Figure 4   Predicted vs. Actual (a) Cumulative and (b) Annual Repeat Transactions

(a)

(b)

Table 5   Expected Number of Repeat Transactions in 2002–2006 as a Function of Recency and Frequency

No. of rpt transactions (1996–2001)	Year of last transaction						
	1995	1996	1997	1998	1999	2000	2001
0	0.07						
1		0.09	0.31	0.59	0.84	1.02	1.15
2			0.12	0.54	1.06	1.44	1.67
3				0.22	1.03	1.80	2.19
4					0.58	2.03	2.71
5						1.81	3.23
6							3.75

100004, with better recency but lower frequency, is expected to make 2.71 transactions over the same period—an increase of nearly 50%. This highlights the critically important role of recency, which can also be seen in the steep growth of the curve in Figure 6(b).

• Although donors 100004 and 111103 have different histories, their recency and frequency numbers are identical ($x = 4$, $t_x = 6$); thus, they have the same conditional expectation. Minor, remote differences in purchase histories are deemed to be irrelevant when making predictions using the BG/BB model.

• A donor who has been completely absent since making his or her initial transaction is expected to make only 0.07 repeat transactions over the next five years. However, although each such donor is not particularly valuable alone, it is important to note, as per Table 2, that over 30% of the entire cohort of donors is in this recency/frequency group. Taken together, these donors are expected to make over 240 transactions over the next five years, making them collectively more valuable than about half of the other recency/frequency groups.

Beyond these specific analyses, Table 5 offers additional insights about the broader interplay between recency and frequency. First, note that for any row (i.e., value of x), the expected number of transactions in the forecast period decreases as we move from right to left (i.e., the less recent the last observed transaction). This is as we would expect, because the longer the hiatus in making a purchase, the more likely it is that the customer is "dead." Looking down the columns, however, we see a somewhat different pattern. We first look at 2001 and note that the conditional expectation is clearly an increasing function of the number of repeat transactions made in the six-year calibration period. Looking at the 1997–2000 columns, though, we note that the numbers first increase, then decrease as the number of repeat transactions made in the six-year calibration period decreases. (A similar pattern is observed in the DERT numbers under the Pareto/NBD model reported in Fader et al. 2005.)

To help understand why this is the case, we use (11) and (17) to compute P(alive in 2002) and the mean

have otherwise seemed. (With reference to Figure 6(a), we see that this conditional expectation overestimates the actual mean (3.53) by only 6%.)

• Donor 100009, who had had a perfect record until the most recent year, is expected to make 1.81 transactions over the next five years. In contrast, donor

Figure 5   Predicted vs. Actual Frequency of Repeat Transactions in 2002–2006

Figure 6 Predicted vs. Actual Conditional Expectations of Repeat Transactions in 2002–2006 as a Function of (a) Frequency and (b) Recency

(a)

(b)
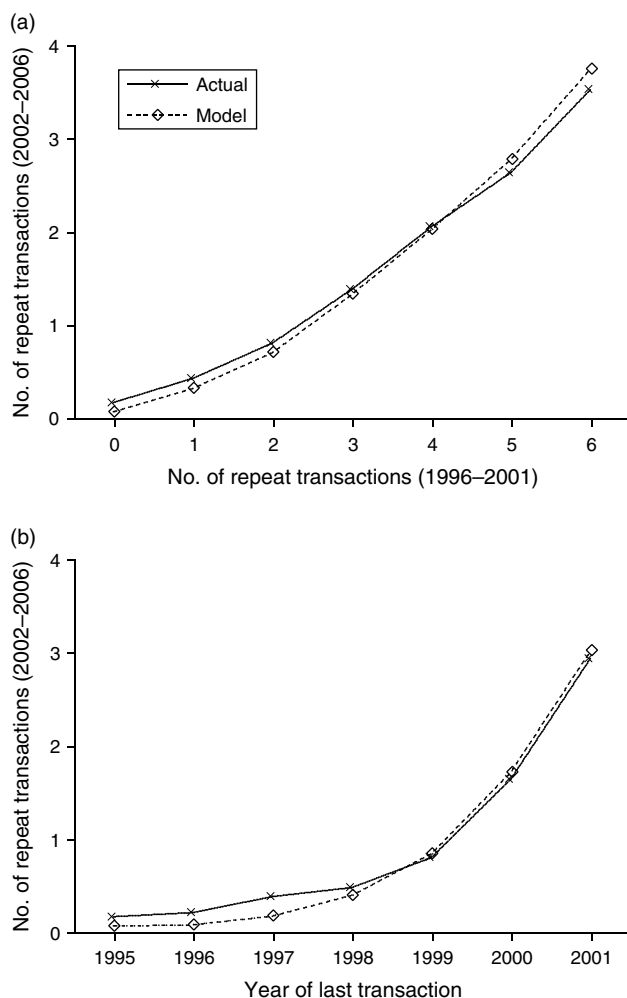
of the marginal posterior distribution of P as a function of recency and frequency. The combinations of the patterns we shall see in these two tables provides an explanation for this somewhat surprising pattern of conditional expectations.

Let us first consider the probability that a customer is alive in 2002; see Table 6. Looking across the columns for any value of x, the observed pattern is as would be expected, with a lower probability of being alive the longer the hiatus in making a donation. Taking a columnwise view, the first thing to note is that all customers who made a transaction in 2001 have the same probability of being alive the following year, regardless of the number of repeat transactions they had prior to that year; this is a natural consequence of the Bernoulli "death" process. Looking at the 1997–2000 columns, we note that the numbers increase as the number of repeat transactions made in the six-year calibration period decreases. The logic behind this is as follows: looking at the 2000 column,

Table 6 P(Alive in 2002) as a Function of Recency and Frequency

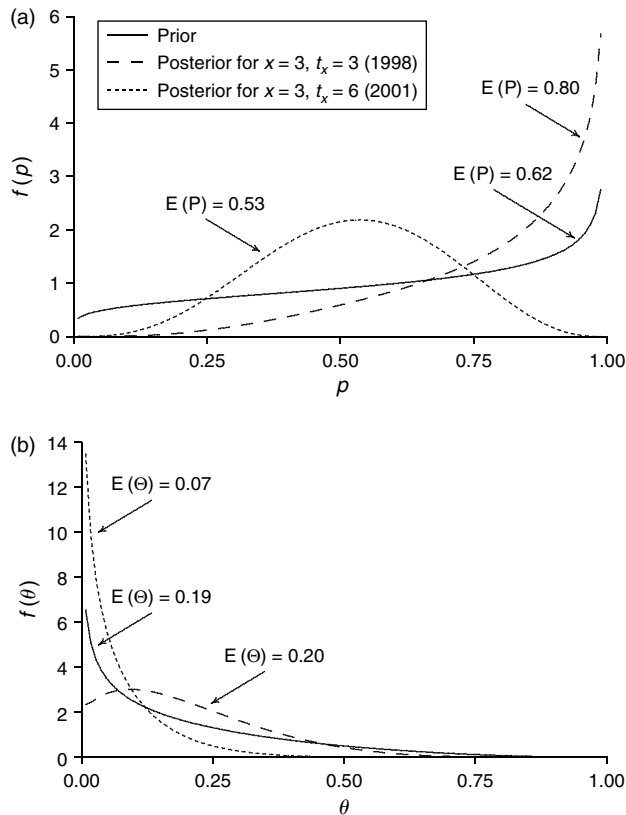| No. of rpt transactions (1996–2001)	Year of last transaction						
	1995	1996	1997	1998	1999	2000	2001
0	0.11						
1		0.07	0.25	0.48	0.68	0.83	0.93
2			0.07	0.30	0.59	0.80	0.93
3				0.10	0.44	0.77	0.93
4					0.20	0.70	0.93
5						0.52	0.93
6							0.93

those customers who made only one repeat transaction will have a lower value of p than those who have made a repeat purchase in all five years, and therefore the fact that no transaction occurred in 2001 can be attributed more to their low probability of making a purchase in any given year than to the possibility of them being dead.

Table 7 reports the mean of the marginal posterior distribution of P. Looking at this table column by column, we see that the posterior mean increases as a function of the number of repeat transactions in the calibration period for any given value of recency. This is intuitive: a smaller number of repeat transactions reflects a lower underlying probability of purchasing at any given transaction opportunity (assuming one is alive). Perhaps less immediately intuitive is the within-row pattern: for a given level of frequency, the underlying probability of purchasing at any given transaction opportunity increases as recency decreases. The reason for this is that, other things being equal, the longer the hiatus since the last transaction, the more likely it is that the customer is dead, and therefore the individual must have had a higher p in order to have the realized number of transactions while alive.

Further insights can be obtained by looking at the marginal posterior distributions of P and Θ, (15) and (16). With reference to Figure 7(a), the prior is the plot of a beta distribution with parameters $\alpha = 1.204$ and $\beta = 0.750$; the overall mean of P across the whole sample is 0.62. With reference to Figure 7(b),

Table 7 Posterior Mean of P as a Function of Recency and Frequency

No. of rpt transactions (1996–2001)	Year of last transaction						
	1995	1996	1997	1998	1999	2000	2001
0	0.49						
1		0.66	0.44	0.34	0.30	0.28	0.28
2			0.75	0.54	0.44	0.41	0.40
3				0.80	0.61	0.54	0.53
4					0.82	0.68	0.65
5						0.83	0.78
6							0.91

Figure 7   Prior and Selected Posterior Distributions of (a) P and (b) Θ

Table 8   Probability of Being Active in 2002–2006 as a Function of Recency and Frequency

No. of rpt transactions (1996–2001)	Year of last transaction						
	1995	1996	1997	1998	1999	2000	2001
0	0.05						
1		0.05	0.17	0.32	0.46	0.56	0.62
2			0.05	0.24	0.48	0.66	0.76
3				0.09	0.40	0.69	0.84
4					0.19	0.66	0.88
5						0.51	0.91
6							0.92

the prior is the plot of a beta distribution with parameters $\gamma = 0.657$ and $\delta = 2.783$; the overall mean of Θ across the whole sample is 0.19. The posterior distribution of P for an individual who made three consecutive repeat purchases with the last one in 1998 has most of its mass to the right; the observed sequence of purchases reflects the high mean of this distribution ($E(P) = 0.80$). At the same time, the three-year hiatus suggests that the supporter is dead as a result of their θ coming from a posterior distribution with an interior mode and with $E(\Theta) = 0.20$.

On the other hand, someone who made three repeat purchases with the last one in 2001 had to be alive over the whole period, which is a result of their θ coming from a beta distribution with most of its mass piled to the left, with $E(\Theta) = 0.07$. The fact that transactions did not occur in three of the six years reflects the fact that their p comes from a distribution with a lower mean ($E(P) = 0.53$).

These relationships between P and Θ suggest that there may be some correlation in the joint posterior distribution (despite the fact we assume independent priors). This is indeed the case, and we explore it with two analyses in Appendix B. (We discuss a model with correlated priors in §5.)

3.1.2. Conditional Penetration. Ever since the publication of Schmittlein et al. (1987), researchers

have shown interest in the P(alive) measure. Although we have reported this quantity as a means of understanding patterns of conditional expectations, we feel that the measure is of limited diagnostic value when viewed by itself. It is a prediction of something that is, by definition, unobservable (i.e., whether or not a customer is still alive at a particular point in time), and thus it is impossible to directly assess its validity. A useful companion measure is a prediction of whether or not the customer will be *active* in the future, that is, whether or not the customer undertakes any transactions in a specified future period of time.[5]

The probability that a customer is active in the 2002–2006 period ($n^* = 5$) is computed as $1 - P(X(n, n + n^*) = 0 \mid x, t_x, n)$ using (12), conditional on each of the 22 (x, t_x) patterns associated with $n = 6$. This conditional penetration is reported in Table 8 as a function of recency (the year of the individual's last transaction) and frequency (the number of repeat transactions).

Comparing Tables 5 and 8, we note that the estimated probabilities of being alive in 2002 are strictly higher than the corresponding conditional 2002–2006 penetration numbers. This makes intuitive sense, but the differences between these measures reflect several factors. First, the P(alive) numbers are just for one year, whereas the penetration numbers are for a five-year period. Second, the mere fact that someone is alive does not mean she will be active, because the latter state depends on the person's underlying transaction probability p. This is very clear when we look at the rightmost column of both tables. Although those people who made a purchase in 2001 have the same probability of being alive, irrespective of frequency, their corresponding probabilities of making at least one transaction in the next five years clearly (and logically) increase as a function of frequency, reflecting

---

[5] Many authors, including Schmittlein et al. (1987), have used the terms "alive" and "active" as synonyms. We feel that this should not be the case, with the term "alive" referring to an unobservable state and the term "active" referring to observable behavior.

in part the associated probabilities of making a purchase at any given transaction opportunity given alive (Table 7). Third, the lower penetration numbers also reflect the fact that inactivity may be due to the person dying in 2003–2006, even if they had been alive in 2002.

In summary, we encourage researchers who might be attracted by the P(alive) measure to also utilize the conditional penetration numbers, because they reflect an observable quantity (i.e., whether or not the customer is active).

3.2. Pooled Analysis

The analyses presented above all focused on a single cohort, the group of individuals who made their first-ever donation during 1995. However, as noted earlier, we have data for a total of six cohorts. At first glance we may be tempted to apply the model cohort by cohort; unfortunately, we are not able to estimate a complete set of cohort-specific parameters. Consider, for instance, the 2000 cohort: we only have one observation per customer—whether or not each new donor made a repeat donation in 2001 (i.e., $n = 1$)—and as such cannot identify the model parameters. The obvious, albeit possibly restrictive, solution is to pool all six cohorts and estimate a single set of model parameters. We now turn our attention to such an analysis, examining how well the BG/BB model predicts the behavior of the complete group of the 56,847 people who made their first-ever donation to the organization between 1995 and 2000.

The maximum-likelihood estimates of the model parameters are reported in Table 9. (Comparing the fit of the BG/BB model with that of the beta-Bernoulli model, we once again note that the addition of the death component results in a major improvement in model fit.) We also note that the BG/BB parameters for the pooled model are remarkably similar to those of the 1995 cohort by itself (Table 4)—this reflects both the high reliability of the model as well as the "poolability" of the cohorts. Figure 8, which compares the expected number of people making $0, 1, \ldots, 6$ repeat transactions between 1996 and 2001 with the observed frequencies, confirms that the model provides a very good fit to the data.

The pooled model continues to accurately track the actual number of repeat transactions over time. Viewing Figure 9(a), which shows the actual versus predicted cumulative number of repeat transactions, we

Figure 8 Predicted vs. Actual Frequency of Repeat Transactions by the 1995–2000 Cohorts
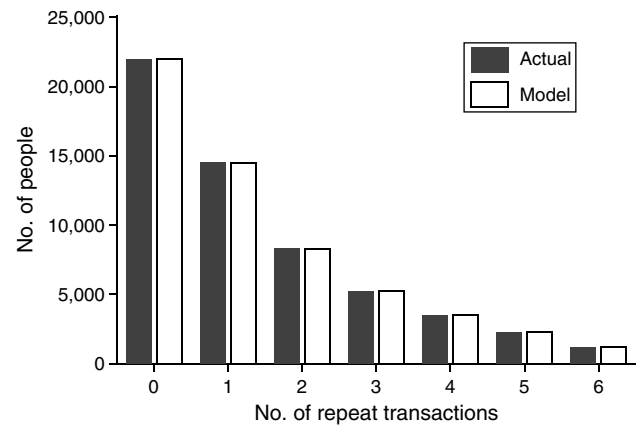
see that the model overforecasts the holdout transactions by a mere 0.25%. Looking at Figure 9(b), which reports these numbers on a year-by-year basis, we note that the BG/BB model clearly captures the underlying trend in repeat transactions. (The repeat transaction numbers rise up to 2001 as new supporters continue to enter the combined pool of donors; after that point, we are focusing on a fixed group of

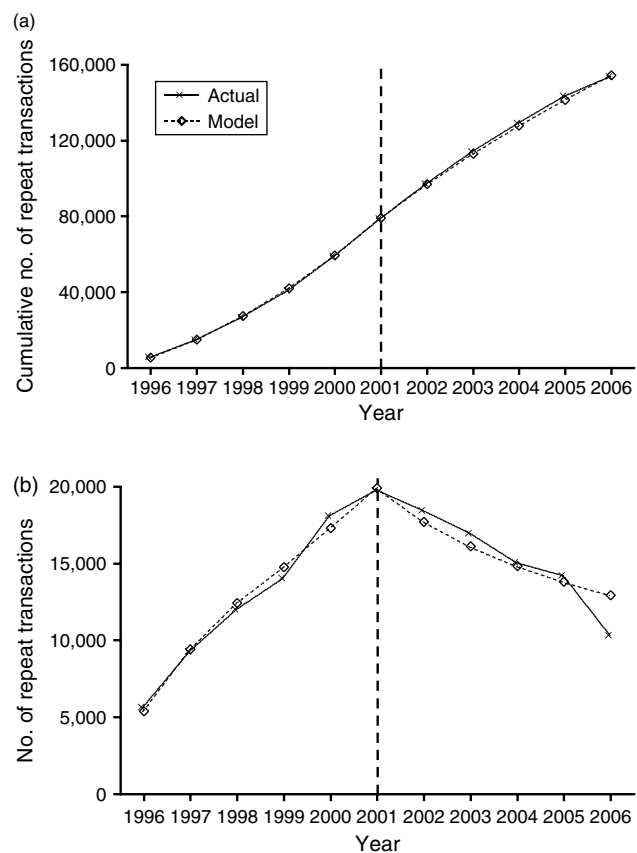**Figure 9** Predicted vs. Actual (a) Cumulative and (b) Annual Repeat Transactions

Table 9 Parameter Estimates, Pooling the 1995–2000 Cohorts

	α	β	γ	δ	LL
BB	0.501	0.753			−115,615.0
BG/BB	1.188	0.749	0.626	2.331	−110,521.0

56,847 potential repeat supporters.) The conditional expectation plots, omitted in the interests of space, are similarly impressive.

This pooled analysis provides a further illustration of the remarkable ability of the BG/BB model to describe and predict the future behavior of a customer base. It is encouraging to see how one set of parameters can capture the behavior of different cohorts acquired across six consecutive years (1995–2000) and project their actions quite accurately into the future.

4. Comparison with the Pareto/NBD Model

Our empirical analysis has focused on the number of repeat transactions. The alert reader will have questioned our use of the term "transactions" because this is not a "necessarily discrete" setting (Figure 1). Strictly speaking, we have been modeling whether or not the supporter has made *any* donation to the organization each year; we have ignored the fact that some supporters may make more than one donation in a given year.

We feel that such an approach is perfectly appropriate for two reasons. First, the majority of the supporter base (71%) made only one donation for each of the years during which a "transaction" occurred. Second, this is the way the nonprofit organization thinks about its donor base; they focus more on whether or not each person has made a donation in any given year (0/1), not as much on the number of donations made. Thus, the 0/1 indicator is the primary behavioral measure recorded in the database provided to us (just as it was for Netzer et al. 2008).

Nevertheless, the fact that 29% of the supporter base made more than one donation in at least one of the years during which a "transaction" occurred may lead some to argue that we should be modelling the number of donations over time rather than annual incidence; the natural model to use for such an approach to the data would be the Pareto/NBD.

Returning to the 1995 cohort, we obtained data on the number of repeat donations made by each supporter within each year (i.e., the binary string characterization of behavior is replaced by a string of nonnegative integers). Given the interval-censored nature of these data, we estimate the parameters of the Pareto/NBD model using the likelihood function given in Fader and Hardie (2005).[6]

The expected number of people making $0, 1, 2, \ldots$ repeat donations between 1996 and 2001 is compared to the actual frequency distribution in Figure 10. In contrast to the performance we normally expect from
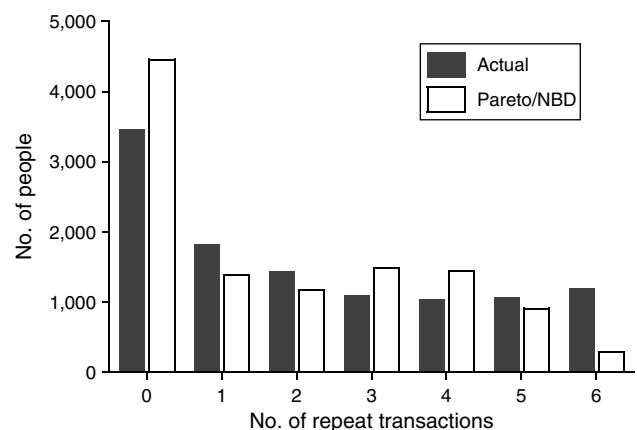
the Pareto/NBD model (e.g., Fader et al. 2005), we note that the Pareto/NBD provides a poor fit to the observed donation data.

Another test of the Pareto/NBD as a model of the donation process is to estimate the implied flow of annual "transactions" (i.e., annual incidence) and then examine how well the model captures and predicts the observed transaction patterns. The expected number of people making $0, 1, \ldots, 6$ repeat transactions between 1996 and 2001 is compared to the actual frequency distribution in Figure 11. In contrast to the fit observed for the BG/BB model in Figure 3, we see that the Pareto/NBD fails to capture the observed annual incidence of donations.

We can also examine how well the model tracks repeat transactions over time, both cumulatively (Figure 12(a)) and year by year (Figure 12(b)). In contrast to the equivalent plots for the BG/BB model (Figures 4(a) and 4(b), respectively), we see that Pareto/NBD fails to track the actual data. The initial

---

[6] The parameter estimates are $\hat{r} = 11.419$, $\hat{\alpha} = 12.865$, $\hat{s} = 0.129$, and $\hat{\beta} = 0.013$, with LL $= -44,506.6$.

Figure 10 Comparing the Number of Repeat Donations as Predicted by the Pareto/NBD Model with the Actual Numbers

Figure 11 Comparing the Number of Repeat Transactions (i.e., Annual Incidence) as Predicted by the Pareto/NBD Model with the Actual Numbers

Figure 12   Predicted vs. Actual (a) Cumulative and (b) Annual Repeat Transactions
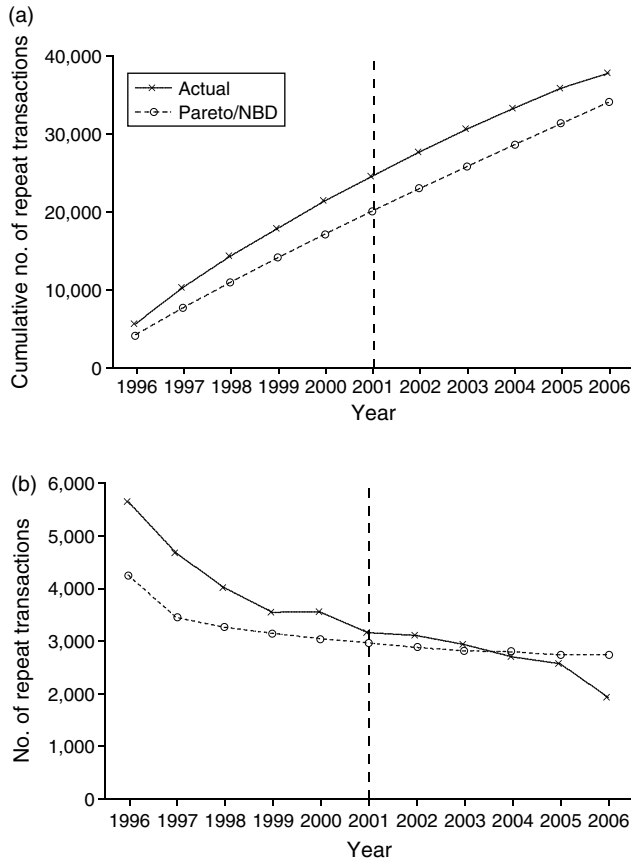
underprediction follows naturally from the overestimation of the number of people making zero donations between 1996 and 2001. We also note that the Pareto/NBD fails to capture the overall rate of decline in transactions over time.

Finally, we examine how well the BG/BB and Pareto/NBD models track (and predict) the evolution of the number of cohort members that ever make a repeat transaction—see Figure 13. Once again we see

Figure 13   Comparing the Number of "Ever-Repeaters" as Predicted by the BG/BB and Pareto/NBD Models with the Actual Number
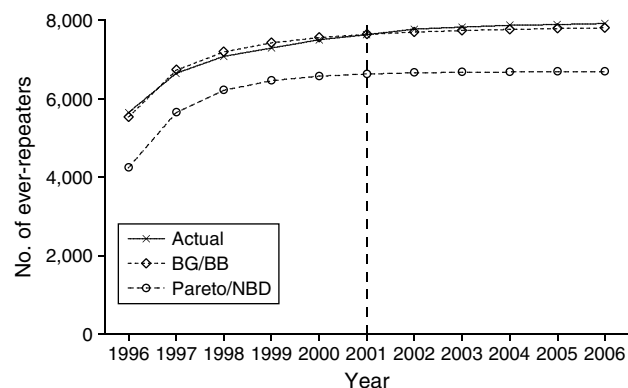
a strong performance by the BG/BB model and a poor performance by the Pareto/NBD model.

To summarize, this analysis has demonstrated that the Pareto/NBD model fails to capture the flow of donations. Treating the data as discrete—even though the underlying process is not "necessarily discrete"—and modeling the flow of transactions (i.e., incidence, rather than the overall number within each discrete time interval) using the BG/BB model is clearly superior.

Why does the Pareto/NBD perform so poorly in this case? The assumption of exponential "interpurchase" times between donations (which yields the Poisson count model) is a dubious one in this setting. Donations are made too "regularly" (e.g., in December of each year) to be accommodated by the "memorylessness" of the exponential/Poisson. Consider, for example, the 1,203 customers who made a donation every year (Table 2). An individual-level Poisson model would take such a high donation rate and (because of its equi-dispersion property) would predict a fairly large number of years with multiple donations. However, each of these customers made, on average, a total of only 1.3 donations per year across the calibration period. The Pareto/NBD simply cannot cope with such a low level of persistent behavior. Schmittlein et al. (1987, p. 17) explicitly acknowledged this limitation as well: "For processes like church attendance and television viewing the opportunities for a transaction occur regularly, so our model is...inappropriate." In contrast, directly modeling *annual incidence*—as opposed to continuous-time purchasing—as a memoryless process (while the customer is alive) is a much more reasonable approach.

5.   Extending the Basic Model
Of all the assumptions associated with the BG/BB model, the one that many readers will have the most problem with is Assumption (6), that the transaction probability p and the dropout probability θ vary independently across customers. This is not nearly as restrictive as it may seem; more formally, we are assuming independent priors, which does not imply independence in the joint posterior distribution of P and Θ. (In fact, we can see some fairly strong correlations in the posterior distributions—see Appendix B.) Nevertheless, we now relax this assumption.

An extremely attractive consequence of Assumptions (4)–(6) (i.e., independent beta-mixing distributions) is that we arrive at simple analytical expressions for all the model quantities of interest, which greatly reduces the barriers to model implementation (e.g., being able to perform all the analysis in an Excel spreadsheet). Ideally, we would like to be able to relax

the independence assumption without losing the ability to derive simple analytical expressions.

The Sarmanov family of distributions, as introduced to marketing by Park and Fader (2004), is a natural starting point, because it allows us to create bivariate distributions with specified marginals. However, a problem with the Sarmanov approach is that the range of its correlation coefficients is narrower than $(-1, 1)$ and is a function of the parameters of the marginal distributions. When we relax Assumption (6) via the bivariate beta distribution used by Danaher and Hardie (2005), we find that the distribution is too constraining (i.e., the estimate of the correlation reaches the limits imposed by the estimated parameters of the marginal beta distributions).

We therefore consider the more flexible S_{BB} distribution (Johnson 1949), also known as the logit-normal distribution; that is,

$$\begin{bmatrix} \text{logit}(p) \\ \text{logit}(\theta) \end{bmatrix} \sim \text{MVN}\left(\begin{bmatrix} \mu_P \\ \mu_\Theta \end{bmatrix}, \begin{bmatrix} \sigma_P^2 & \sigma_{P\Theta} \\ \sigma_{\Theta P} & \sigma_\Theta^2 \end{bmatrix} \right).$$

Because the individual-level process has not changed, the likelihood function for a randomly chosen customer is obtained by taking the expectation of (4) over the joint distribution of P and Θ:

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma} \mid x, t_x, n)$$
$$= \int_0^1 \int_0^1 L(p, \theta \mid x, t_x, n) f(p, \theta \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) \, dp \, d\theta.$$

The major downside of using this distribution is that there is no analytic solution to this double integral. We therefore evaluate the integrals using Monte Carlo simulation; that is, we estimate the model parameters using the method of maximum simulated likelihood (making use of MATLAB). We call this the S_{BB}-G/B model.

We first estimate a constrained version of the model assuming p and θ are assumed to be uncorrelated. With reference to Table 10, we see that model fit is almost identical to that of the original BG/BB model. The associated moments in the (P, Θ) space are also very close to those associated with the BG/BB model. Allowing for a correlation results in a significant improvement in model fit—an increase of 15 log-likelihood points at the cost of one extra parameter. The estimated (prior) correlation between P and Θ is 0.361 (versus the limit of 0.042 associated with using a Sarmanov bivariate beta distribution).

The big question is whether this improvement in model fit leads to any meaningful improvement in the associated predictions. We first consider how well it tracks aggregate repeat transactions over time. The cumulative and year-by-year numbers are plotted in Figure 14. We note that the differences in the predictions associated with the BG/BB and S_{BB}-G/B models

Table 10 Results of the Model That Replaces Independent Beta-Mixing Distributions with an S_{BB} Distribution for Heterogeneity in P and Θ

| | | S_{BB} heterogeneity | |
	BG/BB	Uncorr	Corr
Parameter estimates			
μ_P		0.720	1.119
μ_Θ		−1.993	-2.145
σ_P^2		3.178	3.869
σ_Θ^2		2.219	4.020
$\sigma_{P\Theta}$		—	1.774
LL	−33,225.6	−33,225.7	−33,210.7
Moments in (P, Θ) space			
$E(P)$	0.616	0.614	0.666
$\text{var}(P)$	0.080	0.082	0.084
$E(\Theta)$	0.191	0.189	0.209
$\text{var}(\Theta)$	0.035	0.037	0.058
$\text{corr}(P, \Theta)$	—	—	0.361

are negligible. However, when we look at the distribution of holdout period transactions (Figure 15), it is clear that the S_{BB}-G/B model provides a better prediction of the distribution than the already

Figure 14 Comparing Predicted (a) Cumulative and (b) Annual Repeat Transactions from the BG/BB and S_{BB}-G/B Models vs. Actual
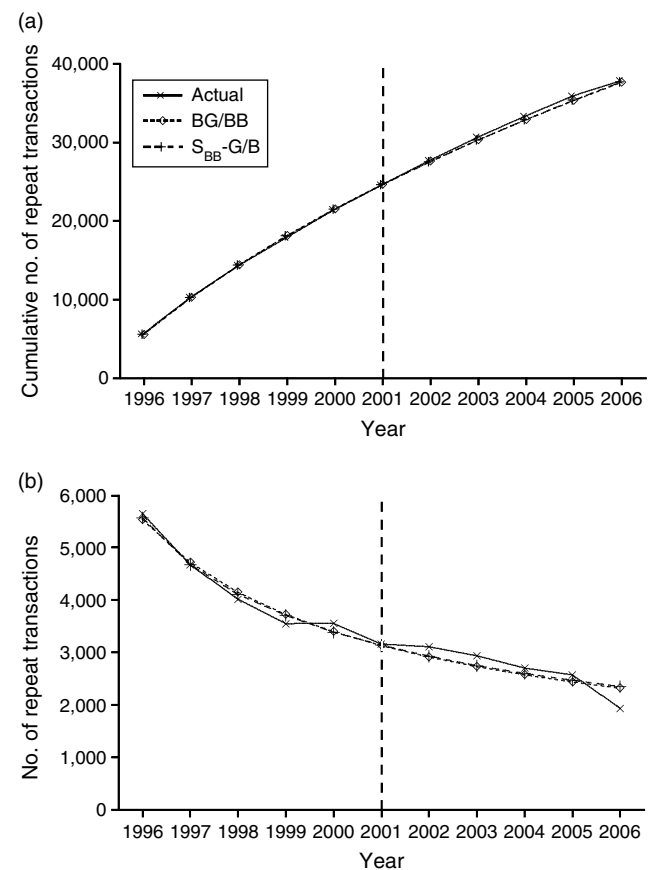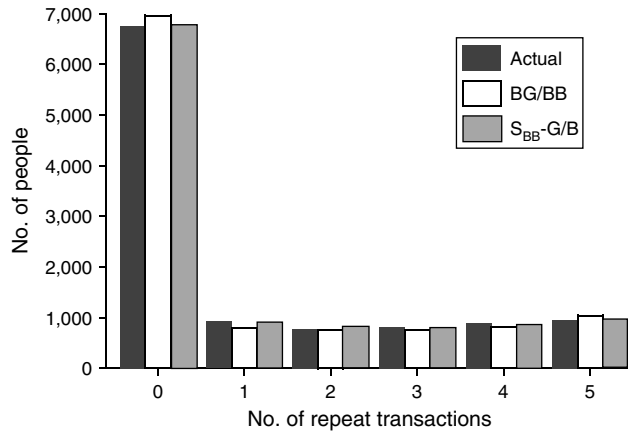
Figure 15   Predicted (from the BG/BB and S_{BB}-G/B Models) vs. Actual Frequency of Repeat Transactions in 2002–2006

Table 11   Expected Number of Repeat Transactions in 2002–2006 as a Function of Recency and Frequency, as Predicted by the S_{BB}-G/B Model

No. of rpt transactions	Year of last transaction						
(1996–2001)	1995	1996	1997	1998	1999	2000	2001
0	0.10						
1		0.10	0.44	0.75	0.93	1.04	1.11
2			0.12	0.66	1.21	1.52	1.68
3				0.22	1.15	1.93	2.24
4					0.56	2.12	2.78
5						1.78	3.26
6							3.59

excellent prediction associated with the BG/BB model.[7]

Turning our attention to the conditional expectations, we first look at the expected number of transactions in the 2002–2006 period ($n^* = 5$) conditional on each of the 22 (x, t_x) patterns associated with $n = 6$. These conditional expectations are reported in Table 11; they are the S_{BB}-G/B model equivalents of the numbers reported in Table 5. We note that these conditional expectations are highly correlated with those associated with the BG/BB model ($r = 0.997$).

To compare these predictions with those associated with the BG/BB model, we report in Figure 16(a) the two sets of conditional expectations along with the average of the number of the transactions that actually occurred in the 2002–2006 forecast period, broken down by the number of repeat transactions in 1996–2001. (As in Figure 6(a), we are averaging over customers with different values of t_x for each x.) Similarly, Figure 16(b) reports the two sets of conditional expectations along with the average of the number of the transactions broken down by the year of the individual's last transaction. (For each t_x, we are averaging over customers with different values of x.) For the most part, the predictions from the two models are very close. Nevertheless, there are some noticeable differences (e.g., a donor who made a repeat transaction every year in the calibration period is expected to make 3.59 transactions over the subsequent five years according to the S_{BB}-G/B model, versus 3.75 under the BG/BB).

In conclusion, we find that, at least for this empirical setting, there is a significant (prior) correlation between the transaction and dropout probabilities; that is, Assumption (6) is violated. However, relaxing

this assumption comes at a cost. Whereas the basic BG/BB model can be implemented in Excel, the S_{BB}-G/B model requires a less accessible computing environment (e.g., MATLAB). Although allowing for this correlation does lead to some improvements in the model's predictive performance, the numbers are sufficiently similar for us to conclude that the cost-benefit

Figure 16   Predicted (from the BG/BB and S_{BB}-G/B Models) vs. Actual Conditional Expectations of Repeat Transactions in 2002–2006 as a Function of (a) Frequency and (b) Recency

(a)

(b)

---
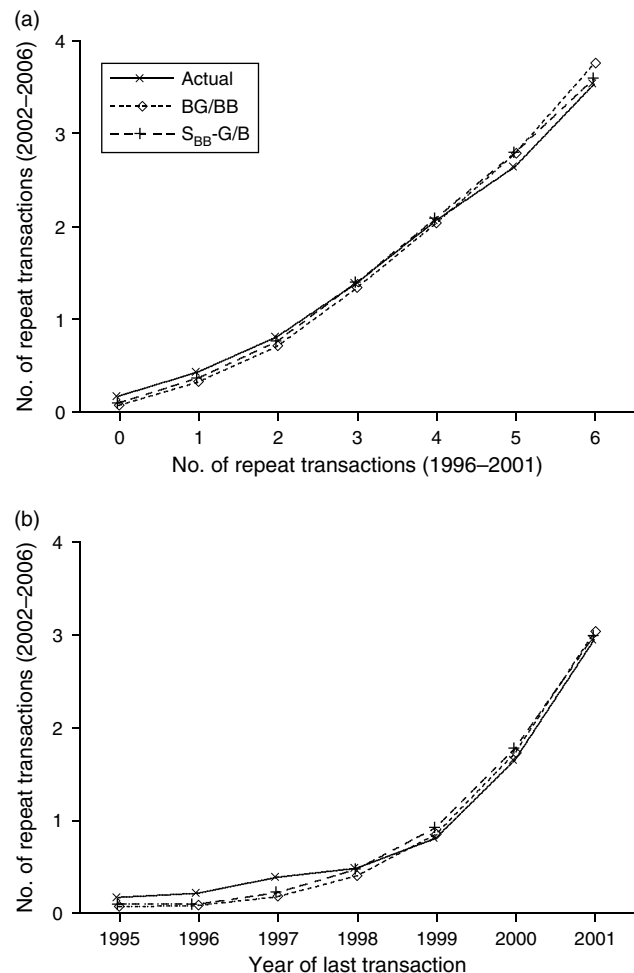
[7] Assessing the relative "fit" using the chi-squared goodness-of-fit measure, we note that it reduces from 47.9 for the BG/BB model to 4.8 for the S_{BB}-G/B model.

trade-off is not immediately obvious. We will revisit this issue in the following section.

6. Discussion

We have developed a new model that can be used to answer standard customer-base analysis questions in noncontractual settings where opportunities for transactions occur at discrete intervals. Using a data set on annual donations made by the supporters of a nonprofit organization located in the midwestern United States, we have demonstrated how the model can be used to compute a number of managerially relevant quantities such as future purchasing patterns, both collectively and individually (conditional on past behavior). In examining these quantities we have observed some interesting effects of past behavior (as summarized by recency and frequency) on predictions about future behavior.

The contractual versus noncontractual distinction that lies at the heart of this work is very similar to Jackson's (1985a, b) "lost-for-good" versus "always-a-share" framework. Rust et al. (2004) observe that such a distinction is important, because the estimates of CLV generated by applying a lost-for-good model to data best characterized by the always-a-share assumption will systematically underestimate true CLV. In a discrete-time always-a-share setting, the BB is the natural benchmark model for purchasing from the firm. However, as shown earlier, it substantially overforecasts cumulative repeat transactions; it fails to capture the "leakage" of customers over time typically observed in an always-a-share setting—also observed by East and Hammond (1996). By allowing for an unobserved death component, the BG/BB can be viewed as a "leaky" version of an always-a-share model.

As we mentioned from the outset of this paper, the BG/BB is the direct analog of the Pareto/NBD as one moves from a continuous-time setting to a discrete-time domain. We have brought up a number of specific examples where this distinction is critically important, as well as some situations (characterized as discretized by recording process in Figure 1) where the analyst might intentionally convert a continuous-time setting into a discrete-time one, primarily to be able to use the BG/BB model instead of the Pareto/NBD. We are aware of several organizations (including hotel chains, financial services firms, and a variety of nonprofits) that have chosen to focus on "discretized" data, either on their own (such as the organization that provided the data used here) or specifically to utilize the BG/BB framework. The fact that they have approached their data management/analysis in such a manner is an indication of the direct applicability of this new model.

Various benefits associated with the BG/BB have been mentioned throughout this paper, and we summarize them here.

• The BG/BB offers tremendous advantages in terms of the required data structures. The size of the data summary required for model estimation is purely a function of the number of transaction opportunities—not the number of customers—and therefore the model is highly "scalable" to customer bases of different sizes. Furthermore, in recognizing that recency and frequency are sufficient summary statistics, the relationship between the number of transaction opportunities and the size of the data set is on the order of n^2, which is a significant reduction compared to using the full binary strings (order 2^n).

• Besides the efficient data requirements, the calculations associated with the model are much simpler than those of the Pareto/NBD. No unconventional or computationally demanding functions are required for parameter estimation or for most of the diagnostic statistics that emerge from the model. Taken together with the aforementioned data advantages, this means that the model is easy to fully implement and utilize within a standard spreadsheet environment, as illustrated in Figure 2. This is very appealing to practitioners, because this reduction in space/effort can be accomplished at virtually no cost (i.e., without sacrificing anything in model performance, as shown in our empirical analyses).

• Pragmatic considerations aside, we see that the Pareto/NBD can fail to capture the flow of donations, be it the actual number or annual incidence. We suspect that there are many settings (particularly when periodic transactions tend to occur during a relatively limited range of time) when these shortcomings of the Pareto/NBD will be quite evident.

• The discrete nature of the data and the associated behavioral "story" lead to model diagnostics that are convenient to display and are readily interpretable. For instance, it is very easy to see and appreciate the nonlinear pattern associated with high frequency and low recency, shown in Table 5. Likewise, a simple examination of that table instantly answers the managerial questions raised in the introduction.

• Finally, it is relatively easy to build and analyze the BG/BB model across multiple cohorts of customers—something that has been done rarely (if ever) in the Pareto/NBD literature. Not only does this make the model even more practical, but the multiyear empirical results shown here offer much stronger support for the model's validity than a single-cohort analysis can provide.

Although the BG/BB is an excellent starting point for modeling discrete-time noncontractual data, there are several natural extensions worth investigating in future research. First, as is the case with the

Pareto/NBD model, the BG/BB model will need to be augmented by a model of purchase amounts when we are interested in the overall monetary value of each customer. A natural candidate would be the gamma-gamma mixture (Colombo and Jiang 1999) that Fader et al. (2005) use in conjunction with the Pareto/NBD model. In situations (such as the data set used here) that are not necessarily discrete and where there is the possibility that more than one transaction could occur in each discrete-time interval, we should derive the monetary-value multiplier by first modeling the number of transactions (conditional on the fact that at least one transaction occurred) and then multiply this by the average value per transaction. A logical model would be the shifted beta-geometric distribution (as used by Morrison and Perry 1970 to model purchase quantity, conditional on purchase incidence).

Second, we may want to allow for a non-zero-order purchasing process at the individual level. A good historical starting point would be the "Brand Loyal Model" (Massy et al. 1970). This would effectively be an extension of the Markov chain model of retail customer behavior at Merrill Lynch by Morrison et al. (1982), an extension in which the "exit parameter" is allowed to be heterogeneous and is estimated directly from the data (as opposed to being derived from other data sources).

The research presented in this paper is clearly anchored in the "probability models for customer-base analysis" tradition, of which the Pareto/NBD is a central model. As Fader and Hardie (2009) note, this stream of research uses combinations of basic probability distributions to develop "simple" models of customer behavior that can be used to make predictions of future behavior conditional on customers' past behavior. It is perhaps useful to reflect on how this fits within the broader customer profitability/CLV/customer equity literature, as exemplified by a number of top managerially oriented books (e.g., Blattberg et al. 2001, Gupta and Lehmann 2005, Kumar 2008, Rust et al. 2000) and the large academic literature (e.g., as reviewed in Blattberg et al. 2008), especially in light of the fact that the effects of factors such as marketing activities are completely ignored.

If one takes an evolutionary model-building view of embedding analytics in an organization (Urban and Karash 1971), models such as the BG/BB represent a natural first step. These models can be implemented by an organization at very low cost. For example, no new software is required and the model can be "coded up" in a blank spreadsheet in a matter of minutes; furthermore, the data requirements are minimal and do not require the merging of databases, as is typically the case when wanting to incorporate the effects

of marketing activities—assuming such data are readily available in the first case.[8] If some of the underlying modeling assumptions are unappealing (e.g., the assumption of independence between the transaction and dropout probabilities), we can create a "version 2.0" of the model that comes at some increased computational cost.

Implicit in these basic models is the assumption that future marketing activities will be basically the same as past marketing activities. The impressive predictive performance of the BG/BB model suggests that this is not an overly restrictive assumption. If there has been some customization of marketing activities on the basis of outputs generated from this model (e.g., after scoring the customer database on the basis of P(alive) or the conditional expectations), then all we would need to do is reestimate the model on an updated data set when it is time to apply the model again in the future. (Given that this can be done in Excel, such reestimation comes at very low cost.) Furthermore, the forecasts generated by the model provide a natural (and low-cost) baseline for examining the performance of the "customized" marketing activities.

Beyond efforts to use the BG/BB for customized marketing activities, a similar iterative approach can be applied to better understand other kinds of time-varying marketing activities. In ongoing field applications of the model, we encourage organizations to rerun the model on a periodic basis to try to detect notable deviations from its baseline predictions, as well as to make inferences about the changing nature of the underlying "buy" and "die" processes. Likewise, we encourage organizations to run the model separately for different "cohorts" of customers, e.g., based on their date and/or channel of acquisition. It is often possible to detect systematic shifts across these incoming customer groups, which can help refine expectations and acquisition tactics for newly acquired customers. Although these efforts admittedly fall short of a full-blown optimization strategy, they help organizations gain a much better feel for the evolving patterns of their customer base and the effectiveness of their marketing efforts.

As this kind of "analytics culture" gets embedded into a marketing organization, we can expect managers to begin to ask deeper kinds of "what-if" and resource allocation questions tied to marketing variables. Assuming all the data are readily available in the organization, it is possible to develop models that incorporate these effects (e.g., Kumar et al. 2008; also see the review by Blattberg et al. 2009). As covariates

---

[8] In the nonprofit example considered in this paper, we know that marketing activities were undertaken but the data were not available. There was no indication that these activities were customized at the donor level.

are incorporated, data structures and model estimation issues become more complex. To the extent that customers have been targeted with different marketing activities on the basis of their past behavior, we must also account for endogeneity. This is clearly a major step up the evolutionary ladder of marketing analytics in the organization. We feel that it is important that any organization embarking on such a journey should learn to walk before they can run, and the BG/BB seems to be a solid way to start the journey.

Acknowledgments

Appendix A. Derivations

In this appendix we present derivations of the key results presented in §2.2. Before starting, we first recall that for $0 < k < 1$,

- The sum of the first n terms of a geometric series is

$$a + ak + ak^2 + \cdots + ak^{n-1} = a\frac{1-k^n}{1-k}, \quad \text{(A1)}$$

- The sum of an infinite geometric series is

$$\sum_{n=0}^{\infty} ak^n = \frac{a}{1-k}, \quad \text{(A2)}$$

and note the following transformation of Euler's integral representation of the Gaussian hypergeometric function ($_2F_1(a, b; c; z)$):

$$\int_0^1 t^{b-1}(1-t)^{c-b-1}(1-zt)^{-a}\,dt$$
$$= B(b, c-b)\,_2F_1(a, b; c; z), \quad c > b. \quad \text{(A3)}$$

A.1. Derivation of (7)

An individual making x purchases had to be alive for at least the first x transactions opportunities. Conditional on p, the probability of observing x transactions out of the i (unobserved) transaction opportunities ($i = x, \ldots, n$) the customer is alive is

$$\binom{i}{x}p^x(1-p)^{i-x}.$$

Removing the conditioning on being alive for i transaction opportunities by multiplying this by the probability that the individual is alive for that length of time gives us

$$P(X(n) = x \mid p, \theta) = \binom{n}{x}p^x(1-p)^{n-x}(1-\theta)^n$$
$$+ \sum_{i=x}^{n-1}\binom{i}{x}p^x(1-p)^{i-x}\theta(1-\theta)^i. \quad \text{(A4)}$$

Taking the expectation of this over the mixing distributions for P and Θ ((1) and (2), respectively) gives us (7).

A.2. Derivation of (8)

Conditional on p and θ, the expected number of transactions over n transaction opportunities is computed as

$$E(X(n) \mid p, \theta) = \sum_{t=1}^{n} P(Y_t = 1 \mid p, \text{alive at } t)P(\text{alive at } t \mid \theta)$$

$$= p\sum_{t=1}^{n}(1-\theta)^t$$

$$= p(1-\theta)\sum_{s=0}^{n-1}(1-\theta)^s,$$

which, recalling (A1) and performing some further algebra,

$$= \frac{p(1-\theta)}{\theta} - \frac{p(1-\theta)^{n+1}}{\theta}. \quad \text{(A5)}$$

Taking the expectation of this over the mixing distributions for P and Θ gives us

$$E(X(n) \mid \alpha, \beta, \gamma, \delta)$$
$$= \left(\frac{\alpha}{\alpha+\beta}\right)\left\{\frac{B(\gamma-1, \delta+1) - \beta(\gamma-1, \delta+n+1)}{B(\gamma, \delta)}\right\}.$$

(Strictly speaking, the use of the integral representation of the beta function to solve the integral associated with taking the expectation over Θ only holds for $\gamma > 1$. However, it can be shown that we arrive at the same result when $0 < \gamma < 1$.) Representing the beta functions in terms of gamma functions and recalling the recursive property of gamma functions gives us (8). Reflecting on the bracketed term in (8) as $n \to \infty$, we note that $E(X(n))$ grows to a limit of

$$\left(\frac{\alpha}{\alpha+\beta}\right)\left(\frac{\delta}{\gamma-1}\right)$$

when $\gamma > 1$. When $\gamma < 1$, there is no limit on $E(X(n))$. (The Pareto/NBD model shares this property regarding the existence of a limit.)

A.3. Derivation of (9) and (10)

Recalling (A4), it follows from the memoryless nature of the death process that

$$P(X(n, n+n^*) = x^* \mid p, \theta, \text{alive at } n)$$
$$= \binom{n^*}{x^*}p^{x^*}(1-p)^{n^*-x^*}(1-\theta)^{n^*}$$
$$+ \sum_{i=x^*}^{n^*-1}\binom{i}{x^*}p^{x^*}(1-p)^{i-x^*}\theta(1-\theta)^i. \quad \text{(A6)}$$

Noting that the probability that someone is alive at n is $(1-\theta)^n$, we have

$$P(X(n, n+n^*) = x^* \mid p, \theta)$$
$$= \delta_{x^*=0}\{1 - (1-\theta)^n\} + \binom{n^*}{x^*}p^{x^*}(1-p)^{n^*-x^*}(1-\theta)^{n+n^*}$$
$$+ \sum_{i=x^*}^{n^*-1}\binom{i}{x^*}p^{x^*}(1-p)^{i-x^*}\theta(1-\theta)^{n+i}.$$

(The first term accounts for the fact that anyone not alive at n will, by definition, not make any purchases in the interval $(n, n+n^*]$.) Taking the expectation of this over the mixing distributions for P and Θ gives us (9).

By definition, $X(n, n+n^*) = X(n+n^*) - X(n)$; it follows that $\mathrm{E}[X(n, n+n^*)] = \mathrm{E}[X(n+n^*)] - \mathrm{E}[X(n)]$. Substituting (8) in this gives us (10).

A.4. Derivation of (11)

Reflecting on (4), the first term is the likelihood of x purchases out of n transaction opportunities under the assumption that the customer was alive for all n transaction opportunities. (The other terms account for the possibility that the individual died before n.) Using Bayes' theorem, it follows that the probability that a customer with purchase history (x, t_x, n) is alive at n is

$$\mathrm{P}(\text{alive at } n \mid p, \theta, x, t_x, n) = \frac{p^x(1-p)^{n-x}(1-\theta)^n}{\mathrm{L}(p, \theta \mid x, t_x, n)}. \quad (A7)$$

It follows that

$$\mathrm{P}(\text{alive at } n+1 \mid p, \theta, x, t_x, n)$$
$$= \frac{p^x(1-p)^{n-x}(1-\theta)^{n+1}}{\mathrm{L}(p, \theta \mid x, t_x, n)}. \quad (A8)$$

By Bayes' theorem, the joint posterior distribution of P and Θ is given by

$$f(p, \theta \mid \alpha, \beta, \gamma, \delta, x, t_x, n)$$
$$= \frac{\mathrm{L}(p, \theta \mid x, t_x, n)f(p \mid \alpha, \beta)f(\theta \mid \gamma, \delta)}{\mathrm{L}(\alpha, \beta, \gamma, \delta \mid x, t_x, n)}, \quad (A9)$$

where the individual elements are given in (1), (2), (4), and (5). Taking the expectation of (A8) over the joint posterior distribution of P and Θ gives us (11).

By the same logic, we can derive an expression for the probability that a customer with purchase history (x, t_x, n) is alive at transaction opportunity $n+m$. Conditional on p and θ,

$$\mathrm{P}(\text{alive at } n+m \mid p, \theta, x, t_x, n) = \frac{p^x(1-p)^{n-x}(1-\theta)^{n+m}}{\mathrm{L}(p, \theta \mid x, t_x, n)}.$$

Taking the expectation of this over the joint posterior distribution of P and Θ yields

$$\mathrm{P}(\text{alive at } n+m \mid \alpha, \beta, \gamma, \delta, x, t_x, n)$$
$$= \frac{\mathrm{B}(\alpha+x, \beta+n-x)}{\mathrm{B}(\alpha, \beta)} \frac{\mathrm{B}(\gamma, \delta+n+m)}{\mathrm{B}(\gamma, \delta)}$$
$$\cdot \mathrm{L}(\alpha, \beta, \gamma, \delta \mid x, t_x, n)^{-1}. \quad (A10)$$

A.5. Derivation of (12)

By definition,

$$\mathrm{P}(X(n, n+n^*) = x^* \mid p, \theta, x, t_x, n)$$
$$= \delta_{x^*=0}\{1 - \mathrm{P}(\text{alive at } n \mid p, \theta, x, t_x, n)\}$$
$$+ \mathrm{P}(X(n, n+n^*) = x^* \mid p, \theta, \text{alive at } n)$$
$$\cdot \mathrm{P}(\text{alive at } n \mid p, \theta, x, t_x, n).$$

Substituting (A6) and (A7) in this, and taking the expectation over the joint posterior distribution of P and Θ, (A9), gives us (12).

A.6. Derivation of (13)

Conditional on p and θ, the expected number of transactions across the next n^* transaction opportunities (i.e., in the interval $(n, n+n^*]$) by a customer with purchase history (x, t_x, n) is

$$\mathrm{E}(X(n, n+n^*) \mid p, \theta, x, t_x, n)$$
$$= \mathrm{E}(X(n, n+n^*) \mid p, \theta, \text{alive at } n)$$
$$\times \mathrm{P}(\text{alive at } n \mid p, \theta, x, t_x, n).$$

Now

$$\mathrm{E}(X(n, n+n^*) \mid p, \theta, \text{alive at } n)$$
$$= \sum_{t=n+1}^{n+n^*} \mathrm{P}(Y_t = 1 \mid p, \text{alive at } t)\mathrm{P}(\text{alive at } t \mid \theta, t > n)$$
$$= p \sum_{t=n+1}^{n+n^*} \frac{(1-\theta)^t}{(1-\theta)^n}$$
$$= p \sum_{s=1}^{n^*} (1-\theta)^s$$
$$= \frac{p(1-\theta)}{\theta} - \frac{p(1-\theta)^{n^*+1}}{\theta}. \quad (A11)$$

Taking the expectation of the product of (A7) and (A11) over the joint posterior distribution of P and Θ, (A9), and simplifying (i.e., representing certain beta functions in terms of gamma functions and exploiting the recursive property of gamma functions) gives us (13).

A.7. Derivation of (14)

The number of discounted expected residual transactions for a customer alive at n is

$$\mathrm{DERT}(d \mid p, \theta, \text{alive at } n)$$
$$= \sum_{t=n+1}^{\infty} \frac{\mathrm{P}(Y_t = 1 \mid p, \text{alive at } t)\mathrm{P}(\text{alive at } t \mid t > n, \theta)}{(1+d)^{t-n}}$$
$$= p \sum_{t=n+1}^{\infty} \frac{(1-\theta)^{t-n}}{(1+d)^{t-n}}$$
$$= p \frac{1-\theta}{1+d} \sum_{s=0}^{\infty} \left(\frac{1-\theta}{1+d}\right)^s,$$

which, recalling (A2),

$$= \frac{p(1-\theta)}{d+\theta}. \quad (A12)$$

Multiplying this by the probability that a customer with purchase history (x, t_x, n) (and latent transaction and dropout probabilities p and θ) is still alive at transaction opportunity n, (A7), gives us

$$\mathrm{DERT}(d \mid p, \theta, x, t_x, n) = \frac{p^{x+1}(1-p)^{n-x}(1-\theta)^{n+1}}{(d+\theta)\mathrm{L}(p, \theta \mid x, t_x, n)}. \quad (A13)$$

Taking the expectation of this over the joint posterior distribution of P and Θ, (A9), gives us

$$\mathrm{DERT}(d \mid \alpha, \beta, \gamma, \delta, x, t_x, n)$$
$$= \mathscr{D} \times \frac{\mathrm{B}(\alpha+x+1, \beta+n-x)}{\mathrm{B}(\alpha, \beta)} \Big/ \mathrm{L}(\alpha, \beta, \gamma, \delta \mid x, t_x, n),$$

where

$$\mathscr{D} = \int_0^1 \frac{1}{d+\theta} \frac{\theta^{\gamma-1}(1-\theta)^{\delta+n}}{B(\gamma,\delta)} \, d\theta,$$

letting $s = 1 - \theta$

$$= \frac{1}{B(\gamma,\delta)} \int_0^1 \frac{1}{(1+d)-s} (1-s)^{\gamma-1} s^{\delta+n} \, ds$$

$$= \frac{1}{B(\gamma,\delta)(1+d)} \int_0^1 s^{\delta+n}(1-s)^{\gamma-1}\left(1-\left(\frac{1}{1+d}\right)s\right)^{-1} ds,$$

which, recalling (A3),

$$= \frac{B(\gamma,\delta+n+1)}{B(\gamma,\delta)(1+d)}\, _2F_1\left(1,\delta+n+1;\gamma+\delta+n+1;\frac{1}{1+d}\right),$$

giving us the expression in (14).

It is interesting to note that this expression for DERT differs from that for the conditional expectation, (13), by a factor of

$$\frac{(\gamma-1)_2F_1(1,\delta+n+1;\gamma+\delta+n+1;1/(1+d))}{(\gamma+\delta+n)(1+d)}$$

$$\cdot \left\{1 - \frac{\Gamma(\gamma+\delta+n)}{\Gamma(1+\delta+n)}\frac{\Gamma(1+\delta+n+n^*)}{\Gamma(\gamma+\delta+n+n^*)}\right\}^{-1}.$$

For any given analysis setting, this is a constant, independent of the customer's exact purchase history. Therefore, any ranking of customers on the basis of DERT will be exactly the same as that derived using the conditional expectation of purchasing over the next n^* periods. When $\gamma > 1$ and $d = 0$ (i.e., there is no discounting of future purchases), this converges to 1 as $n^* \to \infty$.

Because $L(\alpha,\beta,\gamma,\delta \mid x, t_x, n) = 1$ when $x = t_x = n = 0$, it follows that the number of discounted expected transactions (DET) for a just-acquired customer is

$$\mathrm{DET}(d \mid \alpha,\beta,\gamma,\delta)$$

$$= \left(\frac{\alpha}{\alpha+\beta}\right)\left(\frac{\delta}{\gamma+\delta}\right)\frac{_2F_1\left(1,\delta+1;\gamma+\delta+1;1/(1+d)\right)}{1+d}. \quad (A14)$$

To compute DET for a yet-to-be-acquired customer, we need to add 1 to this quantity (i.e., the purchase at time $t = 0$ that corresponds to the customer's first-ever purchase with the firm and therefore starts the transaction opportunity clock).

A.8. Derivation of (15)–(17)

We obtain (15) and (16) by integrating (A9) over θ and p, respectively.

By definition, the (l,m)th product moment $(l, m = 0, 1, 2, \ldots)$ of the joint posterior distribution of P and Θ is

$$E(P^l\Theta^m \mid \alpha,\beta,\gamma,\delta,x,t_x,n)$$

$$= \int_0^1 \int_0^1 p^l\theta^m f(p,\theta \mid \alpha,\beta,\gamma,\delta,x,t_x,n) \, dp \, d\theta,$$

which, recalling (A9),

$$= \int_0^1 \int_0^1 p^l\theta^m \frac{L(p,\theta \mid x,t_x,n)f(p \mid \alpha,\beta)f(\theta \mid \gamma,\delta)}{L(\alpha,\beta,\gamma,\delta \mid x,t_x,n)} \, dp \, d\theta$$

$$= \frac{B(\alpha+l,\beta)}{B(\alpha,\beta)}\frac{B(\gamma+m,\delta)}{B(\gamma,\delta)}$$

$$\times \int_0^1 \int_0^1 \frac{L(p,\theta \mid x,t_x,n)f(p \mid \alpha+l,\beta)f(\theta \mid \gamma+m,\delta)}{L(\alpha,\beta,\gamma,\delta \mid x,t_x,n)} \, dp \, d\theta,$$

which, recalling the derivation of (5), gives us (17).

Appendix B. Correlation Analyses

One of the assumptions associated with the BG/BB model is that the transaction probability p and the dropout probability θ vary independently across customers. At first glance, this may appear to be unrealistic, but it is not nearly as restrictive as it may seem. More formally, we are assuming independent priors, which does not imply independence in the joint posterior distribution of P and Θ; in fact, we can see some fairly strong correlations in the posterior distributions, as we show here in two separate analyses that demonstrate how these correlations can be estimated and interpreted.

First, following an analysis shown in Abe (2009), Figure B.1 is a scatter plot of the means of the marginal posterior distributions of P and Θ. Each circle represents the pair of means for a particular purchase history (x, t_x, n) (computed using (17) with $l = 1, m = 0$ and $l = 0, m = 1$, respectively), and the area of each circle is directly proportional to the number of customers who share the same purchase history (i.e., using the numbers from Table 2). The weighted correlation across the 22 pairs of numbers is -0.42. This implies, as common intuition would suggest, that customers who purchase more frequently (while alive) tend to live longer than light purchasers (but of course we do not want to imply any kind of causal connection here).

However, this analysis tells only part of the story because it only considers the posterior means. When we take into account the full posterior distribution for a given customer, a different correlation analysis emerges. Suppose for each customer in a given recency/frequency group we made a number of draws from their joint posterior distribution—what would be the correlation between p and θ across these draws? The joint posterior distribution of P and Θ is given by (A9). For the special case where $t_x = n$, this collapses to

$$f(p,\theta \mid \alpha,\beta,\gamma,\delta,x,t_x,n) = f(p \mid \alpha+x,\beta+n-x)f(\theta \mid \gamma,\delta+n);$$

Figure B.1   Scatter Plot of the Marginal Posterior Means of P and Θ for the 22 (x,t_x) Patterns Associated with $n = 6$
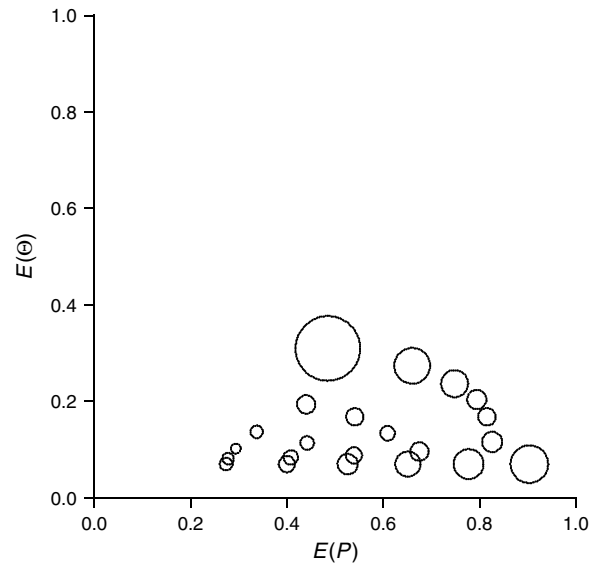
Table B.1 **The Posterior Correlation of P and Θ as a Function of Recency and Frequency**

No. of rpt transactions (1996–2001)	Year of last transaction						
	1995	1996	1997	1998	1999	2000	2001
0	0.258						
1		0.193	0.250	0.203	0.105	0.030	0.000
2			0.165	0.238	0.159	0.047	0.000
3				0.174	0.214	0.071	0.000
4					0.214	0.114	0.000
5						0.190	0.000
6							0.000

i.e., the posterior distribution of P is independent of the posterior distribution of Θ. (Equivalently, the marginal posterior distributions of P and Θ, (15) and (16), collapse to the updated beta distributions $f(p \mid \alpha + x, \beta + n - x)$ and $f(\theta \mid \gamma, \delta + n)$, respectively.) In all other cases, the posterior distribution of an individual's transaction probability is not independent of the posterior distribution of her dropout probability. The joint posterior correlation is given by

$$\text{corr}(P, \Theta \mid \alpha, \beta, \gamma, \delta, x, t_x, n)$$
$$= \frac{\text{E}(P\Theta \mid \cdot) - \text{E}(P \mid \cdot)\text{E}(\Theta \mid \cdot)}{\sqrt{[\text{E}(P^2 \mid \cdot) - \text{E}(P \mid \cdot)^2][\text{E}(\Theta^2 \mid \cdot) - \text{E}(\Theta \mid \cdot)^2]}}, \quad \text{(B1)}$$

where the individual terms are computed using (17). This correlation is reported in Table B.1 as a function of recency (the year of the individual's last transaction) and frequency (the number of repeat transactions).

This table shows that the "intracustomer" correlations are strictly positive (except when $t_x = n$), or, equivalently, if we were to draw from the joint posteriors across all the individuals that are represented within each cell of this table, we would see these positive correlations. In the most extreme case, i.e., when $t_x = n = 0$, we see a fairly strong relationship between p and θ. This makes sense: customers in this cell with a higher purchasing propensity are even more likely (than light purchasers) to be dead. However, across cells, the overall correlation is a fairly strong negative one, as discussed previously. In some sense, this combined analysis (within and across each type of customer) represents a form of Simpson's paradox (Simpson 1951, Wagner 1982).

Taken together, these two analyses provide a more complete picture of the correlations than shown by Abe (2009) and other researchers, who have limited themselves to a simple correlation across the posterior means. More importantly, these analyses put to rest any concerns that a simple empirical Bayesian model with independent priors will be unable to capture and reveal correlations in the underlying processes. To the contrary, these analyses arise quite naturally from the BG/BB model—and the same is true for the Pareto/NBD and other related models.

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.

Berger, P. D., B. Weinberg, R. C. Hanna. 2003. Customer lifetime value determination and strategic implications for a cruise-ship company. *J. Database Marketing Customer Strategy Management* **11**(1) 40–52.

Blattberg, R. C., G. Getz, J. S. Thomas. 2001. *Customer Equity: Building and Managing Relationships as Valuable Assets*. Harvard Business School Publishing, Boston.

Blattberg, R. C., B.-D. Kim, S. A. Neslin. 2008. *Database Marketing: Analyzing and Managing Customers*. Springer, New York.

Blattberg, R. C., E. C. Malthouse, S. A. Neslin. 2009. Customer lifetime value: Empirical generalizations and some conceptual questions. *J. Interactive Marketing* **23**(2) 157–168.

Chatfield, C., G. J. Goodhardt. 1970. The beta-binomial model for consumer purchasing behaviour. *Appl. Statist.* **19**(3) 240–250.

Colombo, R., W. Jiang. 1999. A stochastic RFM model. *J. Interactive Marketing* **13**(3) 2–12.

Danaher, P. J., B. G. S. Hardie. 2005. Bacon with your eggs? Applications of a new bivariate beta-binomial distribution. *Amer. Statistician* **59**(November) 282–286.

East, R., K. Hammond. 1996. The erosion of repeat-purchase loyalty. *Marketing Lett.* **7**(2) 163–171.

Easton, G. 1980. Stochastic models of industrial buying behaviour. *OMEGA* **8**(1) 63–69.

Ehrenberg, A. S. C. 1988. *Repeat-Buying*, 2nd ed. Charles Griffin & Company, London.

Fader, P. S., B. G. S. Hardie. 2005. Implementing the Pareto/NBD model given interval-censored data. Retrieved June 26, 2010, http://brucehardie.com/notes/011/.

Fader, P. S., B. G. S. Hardie. 2009. Probability models for customer-base analysis. *J. Interactive Marketing* **23**(1) 61–69.

Fader, P. S., B. G. S. Hardie, K. L. Lee. 2005. RFM and CLV: Using iso-value curves for customer base analysis. *J. Marketing Res.* **42**(4) 415–430.

Gupta, S., D. R. Lehmann. 2005. *Managing Customers as Investments: The Strategic Value of Customers in the Long Run*. Wharton School Publishing, Upper Saddle River, NJ.

Jackson, B. B. 1985a. Build customer relationships that last. *Harvard Bus. Rev.* **63**(November–December) 120–128.

Jackson, B. B. 1985b. *Winning and Keeping Industrial Customers*. Lexington Books, New York.

Johnson, N. L. 1949. Bivariate distributions based on simple translation systems. *Biometrika* **36**(3–4) 297–304.

Kumar, V. 2008. *Managing Customers for Profit*. Wharton School Publishing, Upper Saddle River, NJ.

Kumar, V., R. Venkatesan, T. Bohling, D. Beckmann. 2008. Practice Prize Report—The power of CLV: Managing customer lifetime value at IBM. *Marketing Sci.* **27**(4) 585–599.

Mason, C. H. 2003. Tuscan lifestyles: Assessing customer lifetime value. *J. Interactive Marketing* **17**(4) 54–60.

Massy, W. F., D. B. Montgomery, D. G. Morrison. 1970. *Stochastic Models of Buying Behavior*. MIT Press, Cambridge, MA.

Morrison, D. G., A. Perry. 1970. Some data based models for analyzing sales fluctuations. *Decision Sci.* **1**(3–4) 258–274.

Morrison, D. G., D. C. Schmittlein. 1988. Generalizing the NBD model for customer purchases: What are the implications and is it worth the effort? *J. Bus. Econom. Statist.* **6**(2) 145–159.

Morrison, D. G., R. D. H. Chen, S. L. Karpis, K. E. A. Britney. 1982. Modelling retail customer behavior at Merrill Lynch. *Marketing Sci.* **1**(2) 123–141.

Netzer, O., J. M. Lattin, V. Srinivasan. 2008. A hidden Markov model of customer relationship dynamics. *Marketing Sci.* **27**(2) 185–204.

Park, Y.-H., P. S. Fader. 2004. Modeling browsing behavior at multiple websites. *Marketing Sci.* **23**(3) 280–303.

Pfeifer, P. E., M. E. Haskins, R. M. Conroy. 2005. Customer lifetime value, customer profitability, and the treatment of acquisition spending. *J. Managerial Issues* **17**(1) 11–25.

Piersma, N., J.-J. Jonker. 2004. Determing the optimal direct frequency. *Eur. J. Oper. Res.* **158**(1) 173–182.

Rosset, S., E. Neumann, U. Eick, N. Vatnik. 2003. Customer lifetime value models for decision support. *Data Mining Knowledge Discovery* **7**(3) 321–339.

Rust, R. T., K. N. Lemon, V. A. Zeithaml. 2004. Return on marketing: Using customer equity to focus marketing strategy. *J. Marketing* **68**(1) 109–127.

Rust, R. T., V. A. Zeithaml, K. N. Lemon. 2000. *Driving Customer Equity*. The Free Press, New York.

Schmittlein, D. C., R. A. Peterson. 1994. Customer base analysis: An industrial purchase process application. *Marketing Sci.* **13**(1) 41–67.

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.

Simpson, E. H. 1951. The interpretation of interaction in contingency tables. *J. Roy. Statist. Soc. Ser. B* **13**(2) 238–241.

Skellam, J. G. 1948. A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials. *J. Roy. Statist. Soc. Ser. B* **10**(2) 257–261.

Urban, G. L., R. Karash. 1971. Evolutionary model building. *J. Marketing Res.* **8**(1) 62–66.

Wagner, C. H. 1982. Simpson's paradox in real life. *Amer. Statistician* **36**(1) 46–48.