"Counting Your Customers" the Easy Way: An Alternative to the Pareto/NBD Model

Peter S. Fader
Bruce G. S. Hardie
Ka Lok Lee[1]

August 2003
Revised April 2004

[1]Peter S. Fader is the Frances and Pei-Yuan Chia Professor of Marketing at the Wharton School of the University of Pennsylvania (address: 749 Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104-6340; phone: 215.898.1132; email: faderp@wharton.upenn.edu; web: www.petefader.com). Bruce G. S. Hardie is Associate Professor of Marketing, London Business School (email: bhardie@london.edu; web: www.brucehardie.com). Ka Lok Lee is a graduate of the Wharton School of the University of Pennsylvania. The first author acknowledges the support of the Wharton-SMU (Singapore Management University) Research Center. The second author acknowledges the support of ESRC grant R000223742 and the London Business School Centre for Marketing.

Abstract

"Counting Your Customers" the Easy Way:
An Alternative to the Pareto/NBD Model

Today's managers are very interested in predicting the future purchasing patterns of their customers, which can then serve as an input into "lifetime value" calculations. Among the models that provide such capabilities, the Pareto/NBD "Counting Your Customers" framework proposed by Schmittlein, Morrison, and Colombo (1987) is highly regarded. But despite the respect it has earned, it has proven to be a difficult model to implement, particularly because of computational challenges associated with parameter estimation.

We develop a new model, the beta-geometric/NBD (BG/NBD), which represents a slight variation in the behavioral "story" associated with the Pareto/NBD, but it is vastly easier to implement. We show, for instance, how its parameters can be obtained quite easily in Microsoft Excel. The two models yield very similar results in a wide variety of purchasing environments, leading us to suggest that the BG/NBD could be viewed as an attractive alternative to the Pareto/NBD in most applications.

Keywords: Customer Base Analysis, Repeat Buying, Pareto/NBD, Probability Models, Forecasting, Lifetime Value

1 Introduction

Faced with a database containing information on the frequency and timing of transactions for a list of customers, it is natural to try to make forecasts about future purchasing. These projections often range from aggregate sales trajectories (e.g., for the next 52 weeks), to individual-level conditional expectations (i.e., the best guess about a particular customer's future purchasing, given information about his past behavior). Many other related issues may arise from a customer-level database, but these are typical of the questions that a manager should initially try to address. This is particularly true for any firm with serious interest in tracking and managing "customer lifetime value" (CLV) on a systematic basis. There is a great deal of interest, among marketing practitioners and academics alike, in developing models to accomplish these tasks.

One of the first models to explicitly address these issues is the Pareto/NBD "Counting Your Customers" framework originally proposed by Schmittlein, Morrison, and Colombo (1987), hereafter SMC. This model describes repeat-buying behavior in settings where customer "dropout" is unobserved: it assumes that customers buy at a steady rate (albeit in a stochastic manner) for a period of time, and then become inactive. More specifically, time to "dropout" is modelled using the Pareto (exponential-gamma mixture) timing model, and repeat-buying behavior while active is modelled using the NBD (Poisson-gamma mixture) counting model. The Pareto/NBD is a powerful model for customer base analysis, but its empirical application can be challenging, especially in terms of parameter estimation.

Perhaps because of these operational difficulties, relatively few researchers actively followed up on the SMC paper soon after it was published (as judged by citation counts). But it has received a steadily increasing amount of attention in recent years as many researchers and managers have become concerned about issues such as customer churn, attrition, retention, and CLV. While a number of researchers (e.g., Balasubramanian et al. 1998; Jain and Singh 2002; Mulhern 1999; Niraj et al. 2001) refer to the applicability and usefulness of the Pareto/NBD, only a small handful claim to have actually implemented it. Nevertheless, some of these papers (e.g., Reinartz and Kumar 2000; Schmittlein and Peterson 1994) have, in turn, become quite popular and widely cited.

The objective of this paper is to develop a new model, the beta-geometric/NBD (BG/NBD), which represents a slight variation in the behavioral "story" that lies at the heart of SMC's original work, but it is vastly easier to implement. We show, for instance, how its parameters can be obtained quite easily in Microsoft Excel, with no appreciable loss in the model's ability to fit or predict customer purchasing patterns. We develop the BG/NBD model from first principles, and present the expressions required for making individual-level statements about future buying behavior. We compare and contrast its performance to that of the Pareto/NBD via a simulation and an illustrative empirical application. The two models yield very similar results, leading us to suggest that the BG/NBD should be viewed as an attractive alternative to the Pareto/NBD model.

Before developing the BG/NBD model, we briefly review the Pareto/NBD model (Section 2). In Section 3 we outline the assumptions of the BG/NBD model, deriving the key expressions at the individual-level, and for a randomly-chosen individual, in Sections 4 and 5 respectively. This is followed by the aforementioned simulation and empirical analysis. We conclude with a discussion of several issues that arise from this work.

2 The Pareto/NBD Model

The Pareto/NBD model is based on five assumptions:

i. While active, the number of transactions made by a customer in a time period of length t is distributed Poisson with transaction rate λ.

ii. Heterogeneity in transaction rates across customers follows a gamma distribution with shape parameter r and scale parameter α.

iii. Each customer has an unobserved "lifetime" of length τ. This point at which the customer becomes inactive is distributed exponential with dropout rate μ.

iv. Heterogeneity in dropout rates across customers follows a gamma distribution with shape parameter s and scale parameter β.

v. The transaction rate λ and the dropout rate μ vary independently across customers.

The Pareto/NBD (and, as we will see shortly, the BG/NBD) requires only two pieces of information about each customer's past purchasing history: his "recency" (when his last transaction occurred) and "frequency" (how many transactions he made in a specified time period). The notation used to represent this information is $(X = x, t_x, T)$, where x is the number of transactions observed in the time period $(0, T]$ and t_x $(0 < t_x \leq T)$ is the time of the last transaction. Using these two key summary statistics, SMC derive expressions for a number of managerially relevant quantities, such as:

- $E[X(t)]$, the expected number of transactions in a time period of length t (SMC, equation 17), which is central to computing the expected transaction volume for the whole customer base over time.
- $P(X(t) = x)$, the probability of observing x transactions in a time period of length t (SMC, equations A40, A43, and A45).
- $E(Y(t) \mid X = x, t_x, T)$, the expected number of transactions in the period $(T, T + t]$ for an individual with observed behavior $(X = x, t_x, T)$ (SMC, equation 22).

The likelihood function associated with the Pareto/NBD model is quite complex, involving numerous evaluations of the Gaussian hypergeometric function. Besides being unfamiliar to most marketing researchers, multiple evaluations of the Gaussian hypergeometric are very demanding from a computational standpoint. Furthermore, the precision of some numerical procedures used to evaluate this function can vary substantially over the parameter space (Lozier and Olver 1995); this can cause major problems for numerical optimization routines as they search for the maximum of the likelihood function.

To the best of our knowledge, the only published paper reporting a successful implementation of the Pareto/NBD model using standard maximum likelihood estimation (MLE) techniques is Reinartz and Kumar (2003), and the authors comment on the associated computational burden. As an alternative to MLE, SMC proposed a three-step method-of-moments estimation procedure, which was further refined by Schmittlein and Peterson (1994). While simpler than MLE, the proposed algorithm is still not easy to implement; furthermore, it does not have the desirable statistical properties commonly associated with MLE. In contrast, the BG/NBD model, to be

introduced in the next section, can be implemented very quickly and efficiently via MLE, and its parameter estimation does not require any specialized software or the evaluation of any unconventional mathematical functions.

3 BG/NBD Assumptions

Most aspects of the BG/NBD model directly mirror those of the Pareto/NBD. The only difference lies in the story being told about how/when customers become inactive. The Pareto timing model assumes that dropout can occur at any point in time, independent of the occurrence of actual purchases. If we assume instead that dropout occurs immediately after a purchase, we can model this process using the beta-geometric (BG) model.

More formally, the BG/NBD model is based on the following five assumptions (the first two of which are identical to the corresponding Pareto/NBD assumptions):

i. While active, the number of transactions made by a customer follows a Poisson process with transaction rate λ. This is equivalent to assuming that the time between transactions is distributed exponential with transaction rate λ, i.e.,

$$f(t_j \,|\, t_{j-1}; \lambda) = \lambda e^{-\lambda(t_j - t_{j-1})}\,, \quad t_j > t_{j-1} \geq 0\,.$$

ii. Heterogeneity in λ follows a gamma distribution with pdf

$$f(\lambda \,|\, r, \alpha) = \frac{\alpha^r \lambda^{r-1} e^{-\lambda\alpha}}{\Gamma(r)}\,, \quad \lambda > 0\,. \tag{1}$$

iii. After any transaction, a customer becomes inactive with probability p. Therefore the point at which the customer "drops out" is distributed across transactions according to a (shifted) geometric distribution with pmf

$$P(\text{inactive immediately after } j\text{th transaction}) = p(1-p)^{j-1}\,, \quad j = 1, 2, 3, \ldots\,.$$

iv. Heterogeneity in p follows a beta distribution with pdf

$$f(p \mid a, b) = \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)}, \quad 0 \le p \le 1\,. \tag{2}$$

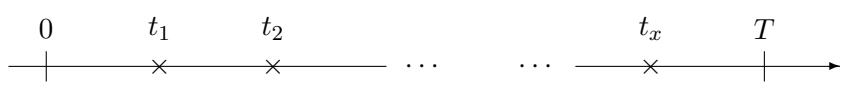
where $B(a,b)$ is the beta function, which can be expressed in terms of gamma functions: $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$.

v. The transaction rate λ and the dropout probability p vary independently across customers.

4 Model Development at the Individual Level

4.1 Derivation of the Likelihood Function

Consider a customer who had x transactions in the period $(0, T]$ with the transactions occurring at $t_1, t_2, \ldots, t_x$:



We derive the individual-level likelihood function in the following manner:

- the likelihood of the first transaction occurring at t_1 is a standard exponential likelihood component, which equals $\lambda e^{-\lambda t_1}$.
- the likelihood of the second transaction occurring at t_2 is the probability of remaining active at t_1 times the standard exponential likelihood component, which equals $(1-p)\lambda e^{-\lambda(t_2-t_1)}$.

  ...
- the likelihood of the xth transaction occurring at t_x is the probability of remaining active at t_{x-1} times the standard exponential likelihood component, which equals $(1-p)\lambda e^{-\lambda(t_x-t_{x-1})}$.
- the likelihood of observing zero purchases in $(t_x, T]$ is the probability the customer became inactive at t_x, plus the probability he remained active but made no purchases in this interval, which equals $p + (1-p)e^{-\lambda(T-t_x)}$.

Therefore,

$$L(\lambda, p \,|\, t_1, t_2, \ldots, t_x, T) = \lambda e^{-\lambda t_1}(1-p)\lambda e^{-\lambda(t_2-t_1)} \cdots (1-p)\lambda e^{-\lambda(t_x-t_{x-1})}\{p + (1-p)e^{-\lambda(T-t_x)}\}$$
$$= p(1-p)^{x-1}\lambda^x e^{-\lambda t_x} + (1-p)^x \lambda^x e^{-\lambda T}$$

As pointed out earlier for the Pareto/NBD, note that information on the timing of the x transactions is not required; a sufficient summary of the customer's purchase history is $(X = x, t_x, T)$.

Similar to SMC, we assume that all customers are active at the beginning of the observation period; therefore the likelihood function for a customer making 0 purchases in the interval $(0, T]$ is the standard exponential survival function:

$$L(\lambda \,|\, X = 0, T) = e^{-\lambda T}$$

Thus we can write the individual-level likelihood function as

$$L(\lambda, p \,|\, X = x, T) = (1-p)^x \lambda^x e^{-\lambda T} + \delta_{x>0}\, p(1-p)^{x-1}\lambda^x e^{-\lambda t_x} \tag{3}$$

where $\delta_{x>0} = 1$ if $x > 0$, 0 otherwise.

4.2 Derivation of $P(X(t) = x)$

Let the random variable $X(t)$ denote the number of transactions occurring in a time period of length t (with a time origin of 0). To derive an expression for the $P(X(t) = x)$, we recall the fundamental relationship between inter-event times and the number of events: $X(t) \geq x \Leftrightarrow T_x \leq t$ where T_x is the random variable denoting the time of the xth transaction. Given our assumption regarding the nature of the dropout process,

$$P(X(t) = x) = P(\text{active after } x\text{th purchase}) \cdot P(T_x \leq t \text{ and } T_{x+1} > t)$$
$$+ \delta_{x>0} \cdot P(\text{becomes inactive after } x\text{th purchase}) \cdot P(T_x \leq t)$$

Given the assumption that the time between transactions is characterized by the exponential distribution, $P(T_x \leq t \text{ and } T_{x+1} > t)$ is simply the Poisson probability that $X(t) = x$, and $P(T_x \leq t)$ is the Erlang-x cdf. Therefore

$$P(X(t) = x|\lambda, p) = (1-p)^x \frac{(\lambda t)^x e^{-\lambda t}}{x!} + \delta_{x>0}\, p(1-p)^{x-1}\left[1 - e^{-\lambda t}\sum_{j=0}^{x-1}\frac{(\lambda t)^j}{j!}\right] \tag{4}$$

4.3 Derivation of $E[X(t)]$

Given that the number of transactions follows a Poisson process, $E[X(t)]$ is simply λt if the customer is active at t. For a customer who becomes inactive at $\tau \leq t$, the expected number of transactions in the period $(0, \tau]$ is $\lambda\tau$.

But what is the likelihood that a customer becomes inactive at τ? Conditional on λ and p,

$$\begin{aligned} P(\tau > t) = P(\text{active at } t \,|\, \lambda, p) &= \sum_{j=0}^{\infty}(1-p)^j \frac{(\lambda t)^j e^{-\lambda t}}{j!} \\ &= e^{-\lambda p t} \end{aligned}$$

This implies that the pdf of the dropout time is given by $g(\tau \,|\, \lambda, p) = \lambda p e^{-\lambda p \tau}$. (Note that this takes on an exponential form. But it features an explicit association with the transaction rate λ, in contrast with the Pareto/NBD, which has an exponential dropout process that is independent of the transaction rate.) It follows that the expected number of transactions in a time period of length t is given by

$$\begin{aligned} E(X(t) \,|\, \lambda, p) &= \lambda t \cdot P(\tau > t) + \int_0^t \lambda\tau g(\tau \,|\, \lambda, p) d\tau \\ &= \frac{1}{p} - \frac{1}{p}e^{-\lambda p t} \end{aligned} \tag{5}$$

5 Model Development for a Randomly-Chosen Individual

All the expressions developed above are conditional on the transaction rate λ and the dropout probability p, both of which are unobserved. To derive the equivalent expressions for a ran-

domly chosen customer, we take the expectation of the individual-level result over the mixing distributions for λ and p, as given in (1) and (2). This yields the follow results.

- Taking the expectation of (3) over the distribution of λ and p results in the following expression for the likelihood function for a randomly-chosen customer with purchase history $(X = x, t_x, T)$:

$$L(r, \alpha, a, b \,|\, X = x, t_x, T) = \frac{B(a, b+x)}{B(a,b)} \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)(\alpha+T)^{r+x}} + \delta_{x>0} \frac{B(a+1, b+x-1)}{B(a,b)} \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)(\alpha+t_x)^{r+x}} \tag{6}$$

  The four BG/NBD model parameters (r, α, a, b) can be estimated via the method of maximum likelihood in the following manner. Suppose we have a sample of N customers, where customer i had $X_i = x_i$ transactions in the period $(0, T_i]$, with the last transaction occurring at t_{x_i}. The sample log-likelihood function given by is

$$LL(r, \alpha, a, b) = \sum_{i=1}^{N} \ln \left[L(r, \alpha, a, b \,|\, X_i = x_i, t_{x_i}, T_i) \right] \tag{7}$$

  This can be maximized using standard numerical optimization routines.

- Taking the expectation of (4) over the distribution of λ and p results in the following expression for the probability of observing x purchases in a time period of length t:

$$P(X(t) = x | r, \alpha, a, b) = \frac{B(a, b+x)}{B(a,b)} \frac{\Gamma(r+x)}{\Gamma(r)x!} \left(\frac{\alpha}{\alpha+t}\right)^r \left(\frac{t}{\alpha+t}\right)^x + \delta_{x>0} \frac{B(a+1, b+x-1)}{B(a,b)} \left[1 - \left(\frac{\alpha}{\alpha+t}\right)^r \left\{ \sum_{j=0}^{x-1} \frac{(\Gamma(r+j)}{\Gamma(r)j!} \left(\frac{t}{\alpha+t}\right)^j \right\} \right] \tag{8}$$

- Finally, taking the expectation of (5) over the distribution of λ and p results in the following expression for the expected number of purchases in a time period of length t:

$$E(X(t) \,|\, r, \alpha, a, b) = \frac{a+b-1}{a-1} \left[1 - \left(\frac{\alpha}{\alpha+t}\right)^r {}_2F_1\big(r, b; a+b-1; \tfrac{t}{\alpha+t}\big) \right] \tag{9}$$

  where ${}_2F_1(\cdot)$ is the Gaussian hypergeometric function. (See the Appendix for details of the derivation.)

Note that this final expression requires a single evaluation of the Gaussian hypergeometric function. But it is important to emphasize that this expectation is only used *after* the likelihood function has been maximized. A single evaluation of the Gaussian hypergeometric function for a given set of parameters is relatively straightforward, and can be closely approximated with a polynomial series, even in a modeling environment such as Microsoft Excel.

In order for the BG/NBD model to be of use in a forward-looking customer-base analysis, we need to obtain an expression for the expected number of transactions in a future period of length t for an individual with past observed behavior $(X = x, t_x, T)$. We provide a careful derivation in the Appendix, but here is the key expression:

$$E(Y(t) \,|\, X = x, t_x, T, r, \alpha, a, b) = \frac{\dfrac{a+b+x-1}{a-1}\left[1 - \left(\dfrac{\alpha+T}{\alpha+T+t}\right)^{r+x} {}_2F_1\big(r+x, b+x; a+b+x-1; \tfrac{t}{\alpha+T+t}\big)\right]}{1 + \delta_{x>0}\dfrac{a}{b+x-1}\left(\dfrac{\alpha+T}{\alpha+t_x}\right)^{r+x}} \tag{10}$$

Once again, this expectation requires a single evaluation of the Gaussian hypergeometric function for any customer of interest, but this is not a burdensome task. The remainder of the expression is simple arithmetic.

6 Simulation

While the underlying behavioral story associated with the proposed BG/NBD model is quite similar to that of the Pareto/NBD, we have not yet provided any assurance that the empirical performance of the two models will be closely aligned with each other. In this section, therefore, we discuss a comprehensive simulation study that provides a thorough understanding of when the BG/NBD can (and cannot) serve as a close proxy to the Pareto/NBD. More specifically, we create a wide variety of purchasing environments (by manipulating the four parameters of the Pareto/NBD model) to look for limiting conditions under which the BG/NBD model does a poor job of capturing the underlying purchasing process.

6.1 Simulation Design

To create these simulated purchasing environments, we chose three levels for each of the four Pareto/NBD parameters, then generated a full-factorial design of $3^4 = 81$ different "worlds". For the two shape parameters (r and alpha) we used values of 0.25, 0.50, and 0.75; for each of the two scale parameters (s and beta) we used values of 5, 10, and 15. When we translate these various combinations into meaningful summary statistics it becomes easy to see the wide variation across these simulated worlds. For instance buyer penetration (i.e., the number of customers who make at least one purchase, or $1 - P(0)$) varies from a low of 13% to a high of 76%. Likewise, average purchase frequency (i.e., mean number of purchases among buyers, or $E[X]/(1 - P(0))$) ranges from 2.1 up to 8.2 purchases per period. It is worth noting that this broad range covers the observed values from the original Schmittlein and Peterson (1994) application as well as the actual dataset used in our empirical analysis (to be discussed in the next section).

For each of the 81 simulated worlds, we created a synthetic panel of 4000 households, then simulated the Pareto/NBD purchase (and dropout) process for a period of 104 weeks. We then ran the BG/NBD model on the first 52 weeks for each of these datasets, and used the estimated parameters to generate forecasts for a holdout period covering the remaining 52 weeks. We evaluate the performance of the BG/NBD based on the mean absolute percent error (MAPE) calculated across this 52-week forecast sales trajectory. If the MAPE value is a low number (below, say, 5%), we have faith in the applicability of the BG/NBD for that particular set of underlying parameters; otherwise we need to look more carefully to understand why the BG/NBD is not doing an adequate job of matching the Pareto/NBD sales projection.

6.2 Simulation Results

In general, the BG/NBD performed quite well in this holdout forecasting task. The average value of the MAPE statistic was 2.68%, and the worst case across all 81 worlds was a reasonably acceptable 6.97%. But upon closer inspection we noticed an interesting, systematic trend across the worlds with relatively high values of MAPE. In Table 1 we summarize the relevant summary statistics for the worst ten simulated worlds in contrast with the remaining 71 worlds. Notice that

the BG/NBD forecasts tend to be relatively poor when penetration and/or purchase frequency are extremely low.

	MAPE	Penetration	Avg. Purch. Freq
Worst 10 worlds	5.73	18.8%	1.2
Other 71 worlds	2.26	44.5%	3.9

Table 1: Summary of Simulation Results

Upon further reflection about the differences between the two model structures, this result makes sense. Under the Pareto/NBD model, dropout can occur at any time — even before a customer has made his first purchase after the start of the observation period. But under BG/NBD, a customer can not become inactive before making his first purchase. If penetrations and/or buying rates are fairly high, then this difference becomes relatively inconsequential. But in a world where active buyers are either uncommon or very slow in making their purchases, the Pareto/NBD will tend to outperform the BG/NBD.

Beyond this one source of deviation, there do not appear to be any other patterns associated with higher versus lower values of MAPE. For instance, the Pearson correlation between MAPE and penetration for the 71 worlds with "good behavior" is a modest 0.142. (In contrast, across all 81 worlds, this correlation is 0.379.) So when we set aside the worlds with sparse buying, the BG/NBD appears to be very robust.

It would be a simple matter to extend the BG/NBD model to allow for a segment of "hard core non-buyers". This would require only one additional parameter and would likely overcome this problem completely. But we do not see the likelihood or severity of this problem to be extreme enough to warrant such an extension as part of the basic model. Nevertheless, we encourage managers to continually monitor summary statistics such as penetration and purchase frequency; for many firms this is already a routine practice.

Having established the robustness (and an important limiting condition) about the BG/NBD, we now turn to a more thorough investigation of its performance (relative to the Pareto/NBD) in an actual dataset.

7 Empirical Analysis

We explore the performance of BG/NBD model using data on the purchasing of CDs at the online retailer CDNOW. The full dataset focuses on a single cohort of new customers who made their first purchase at the CDNOW web site in the first quarter of 1997. We have data covering their initial (trial) and subsequent (repeat) purchase occasions for the period January 1997 through June 1998, during which the 23,570 Q1/97 triers bought nearly 163,000 CDs after their initial purchase occasions. (See Fader and Hardie (2001) for further details about this dataset.)

For the purposes of this analysis, we take a 1/10th systematic sample of the customers. We calibrate the model using the repeat transaction data for the 2357 sampled customers over the first half of the 78-week period and forecast their future purchasing over the remaining 39 weeks. For customer i ($i = 1, ..., 2357$), we know the length of the time period during which repeat transactions could have occurred ($T_i = 39 -$ time of first purchase), the number of repeat transactions in this period (x_i) and the time of his last repeat transaction (t_{x_i}). (If $x_i = 0$, $t_{x_i} = 0$.) In contrast to Fader and Hardie (2001), we are focusing on the number of transactions, not the number of CDs purchased.

Maximum likelihood estimates of the model parameters (r, α, a, b) are obtained by maximizing the log-likelihood function given in (7) above. Standard numerical optimization methods are employed, using the Solver tool in Microsoft Excel, to obtain the parameter estimates. (Identical estimates are obtained using the more sophisticated MATLAB programming language.) To implement the model in Excel, we rewrite the log-likelihood function, (6), as

$$L(r, \alpha, a, b \,|\, X = x, t_x, T) = \mathsf{A}_1 \cdot \mathsf{A}_2 \cdot (\mathsf{A}_3 + \delta_{x>0}\, \mathsf{A}_4)$$

where

$$\mathsf{A}_1 = \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)} \qquad \mathsf{A}_2 = \frac{\Gamma(a+b)\Gamma(b+x)}{\Gamma(b)\Gamma(a+b+x)}$$
$$\mathsf{A}_3 = \left(\frac{1}{\alpha+T}\right)^{r+x} \qquad \mathsf{A}_4 = \left(\frac{a}{b+x-1}\right)\left(\frac{1}{\alpha+t_x}\right)^{r+x}$$

This is very easy to code in Excel — see Figure 1 for complete details. (A note on how to implement the model in Excel, along with a copy of the complete spreadsheet, can be found at `<insert URL>`.

	A	B	C	D	E	F	G	H	I
1	r	0.243							
2	alpha	4.414							
3	a	0.793							
4	b	2.426							
5	LL	-9582.4							
6									
7	ID	x	t_x	T	ln(.)	ln(A_1)	ln(A_2)	ln(A_3)	ln(A_4)
8	0001	2	30.43	38.86	-9.4596	-0.8390	-0.4910	-8.4489	-9.4265
9	0002	1	1.71	38.86	-4.4711	-1.0562	-0.2828	-4.6814	-3.3709
10			0.00	38.86	-0.5538	0.3602	0.0000	-0.9140	0.0000
11	0004	0	0.00	38.86	-0.5538	0.3602	0.0000	-0.9140	0.0000
12									
13									
14	0007	1	5.00	38.86					
15	0008	0	0.00	38.86	-0.5538	0.3602	0.0000	-0.9140	0.0000
16	0009	2	35.71	38.86	-9.5367	-0.8390	-0.4910	-8.4489	-9.7432
17	0010	0	0.00	38.86	-0.5538	0.3602	0.0000	-0.9140	0.0000
2362	2355	0	0.00	27.00	-0.4761	0.3602	0.0000	-0.8363	0.0000
2363	2356	4	26.57	27.00	-14.1284	1.1450	-0.7922	-14.6252	-16.4902
2364	2357	0	0.00	27.00	-0.4761	0.3602	0.0000	-0.8363	0.0000

=GAMMALN(B$1+B8)-GAMMALN(B$1)+B$1*LN(B$2)

=IF(B8>0,LN(B$3)-LN(B$4+B8-1)-(B$1+B8)*LN(B$2+C8),0)

=-(B$1+B8)*LN(B$2+D8)

=SUM(E8:E2364)

=F8+G8+LN(EXP(H8)+(B8>0)*EXP(I8))

=GAMMALN(B$3+B$4)+GAMMALN(B$4+B8)-GAMMALN(B$4)-GAMMALN(B$3+B$4+B8)

Figure 1: Screenshot of Excel Worksheet for Parameter Estimation

The parameters of the Pareto/NBD model are also obtained via MLE, but this task could be performed only in MATLAB due to the computational demands of the model. The parameter estimates and corresponding log-likelihood function values for the two models are reported in Table 2. Looking at the log-likelihood function values, we observe that the BG/NBD model provides a better fit to the data.

	BG/NBD	Pareto/NBD
r	0.243	0.553
α	4.414	10.578
a	0.793	
b	2.426	
s		0.606
β		11.669
LL	-9582.4	-9595.0

Table 2: Model Estimation Results

In Figure 2, we examine the fit of these models visually: the expected numbers of people making 0, 1, ..., 7+ repeat purchases in the 39-week model calibration period from the two models are compared to the actual frequency distribution. The fits of the two models are very close. On the basis of the chi-square goodness-of-fit test, we note that the BG/NBD model provides a better fit to the data ($\chi^2_3 = 4.82$, $p = 0.19$) than the Pareto/NBD, $\chi^2_3 = 11.99$, ($p = 0.007$).

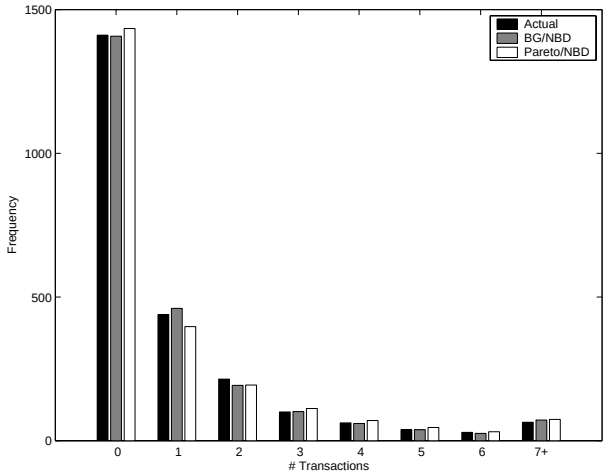


Figure 2: Predicted versus Actual Frequency of Repeat Transactions

The performance of these models becomes more apparent when we consider how well the models track the actual number of (total) repeat transactions over time. During the 39-week calibration period, the tracking performance of the BG/NBD and Pareto/NBD models is practically identical. In the subsequent 39-week forecast period, both models track the actual (cumulative) sales trajectory, with the Pareto/NBD performing slightly better than the BG/NBD (under-forecasting by 2% versus 4%), but both models demonstrate superb tracking/forecasting capabilities.

Our final — and perhaps most critical — examination of the relative performance of the two models focuses on the quality of the predictions of individual-level transactions in the forecast period (Weeks 40–78) conditional on the number of observed transactions in the model calibration period. For the BG/NBD model, these are computed using (10). For the Pareto/NBD, as noted earlier, the equivalent expression is represented by equation (22) in SMC.

In Figure 3, we report these conditional expectations along with the average of the actual number of transactions that took place in the forecast period, broken down by the number of calibration-period repeat transactions. (For each x, we are averaging over customers with different values of t_x.)

Both the BG/NBD and Pareto/NBD models provide excellent predictions of the expected number of transactions in the holdout period. It appears that the Pareto/NBD offers slightly better predictions than the BG/NBD, but it is important to keep in mind that the groups towards

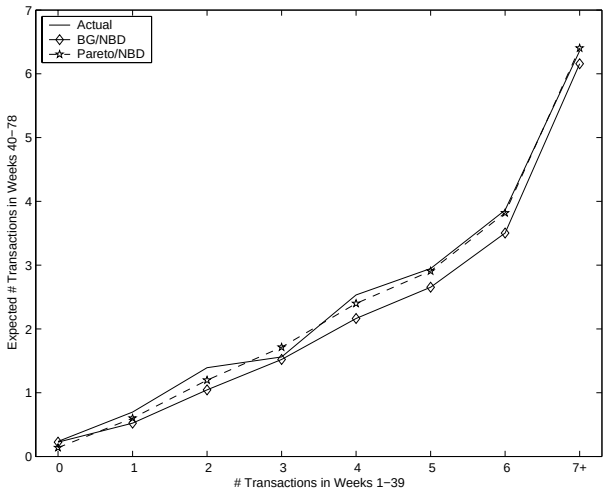


Figure 3: Conditional Expectations

the right of the figure (i.e., buyers with larger values of x in the calibration period) are extremely small. An important aspect that is hard to discern from the figure is the relative performance for the very large "zero class" (i.e., the 1411 people who made no repeat purchases in the first 39 weeks). This group makes a total of 334 transactions in weeks 40–78, which comprises 18% of all of the forecast period transactions. (This is second only to the 7+ group, which accounts for 22% of the forecast period transactions.) The BG/NBD conditional expectation for the zero class is 0.23, which is much closer to the actual average (334/1411=0.24) than that predicted by the Pareto/NBD (0.14).

Nevertheless, these differences are not necessarily meaningful. Taken as a whole across the full set of 2357 customers, the predictions for the BG/NBD and Pareto/NBD models are indistinguishable from each other and from the actual transaction numbers. This is confirmed by a three-group ANOVA ($F_{2,7068} = 2.65$), which is not significant at the usual 5% level. This analysis demonstrates the high degree of validity of both models, particularly for the purposes of forecasting a customer's future purchasing, conditional on his past buying behavior.

8 Discussion

Many researchers have praised the Pareto/NBD model for its sensible behavioral story, its excellent empirical performance, and the useful managerial diagnostics that arise quite naturally

from its formulation. We fully agree with these positive assessments and have no misgivings about the model whatsoever, besides its computational complexity. It is simply our intention to make this type of modeling framework more broadly accessible so that many researchers and practitioners can benefit from the original ideas of SMC.

The BG/NBD model arises by making a small, relatively inconsequential, change to the Pareto/NBD assumptions. The transition from an exponential distribution to a geometric process (to capture customer dropout) does not require any different psychological theories nor does it have any noteworthy managerial implications. When we evaluate the two models on their primary outcomes (i.e., their ability to fit and predict repeat transaction behavior), they are effectively indistinguishable from each other.

As Albers (2000) notes, the use of marketing models in actual practice is becoming less of an exception, and more of a rule, because of spreadsheet software. It is our hope that the ease with which the BG/NBD model can be implemented in a familiar modeling environment will encourage more firms to take better advantage of the information already contained in their customer transaction databases. Furthermore, as key personnel become comfortable with this type of model, we can expect to see growing demand for more complete (and complex) models — and more willingness to commit resources to them.

Beyond the purely technical aspects involved in deriving the BG/NBD model and comparing it to the Pareto/NBD, we have attempted to highlight some important managerial aspects associated with this kind of modeling exercise. For instance, to the best of our knowledge, this is only the second empirical validation of the Pareto/NBD model — the first being Schmittlein and Peterson (1994). (Other researchers (e.g., Reinartz and Kumar 2000, 2003; Wu and Chen 2000) have employed the model extensively, but do not report on its performance in a holdout period.) We find that both models yield very accurate forecasts of future purchasing, both at the aggregate level as well as at the level of the individual (conditional on past purchasing).

Besides using these empirical tests as a basis to compare models, we also want to call more attention to these analyses — with particular emphasis on conditional expectations — as the proper yardsticks that all researchers should use when judging the absolute performance of other forecasting models for CLV-related applications. It is important for a model to be able to

accurately project the future purchasing behavior of a broad range of past customers, and its performance for the zero class is especially critical, given the typical size of that "silent" group.

In using this model, there are several implementation issues to consider. First, the model should be applied separately to customer cohorts defined by the time (e.g., quarter) of acquisition, acquisition channel, etc. (Blattberg et al. 2001). (For a very mature customer base, the model could be applied to coarse RFM-based segments.) Second, if we are using one cohort's parameters as the basis for, say, another cohort's conditional expectation calculations, we must be confident that the two cohorts are comparable. Third, we must acknowledge an implicit assumption when using the forecasts generated using a model such as that developed in the paper: we are assuming that future marketing activities targeted at the group of customers will basically be the same as those observed in the past. (Of course, such models can be used to provide a baseline against which we can examine the impact of changes in marketing activity.) Finally, as with the Pareto/NBD, the BG/NBD must be augmented by a model of purchase amount before it can be used as the basis for CLV calculations. Two candidate models are the normal-normal mixture (Schmittlein and Peterson 1994) and the gamma-gamma mixture (Colombo and Jiang 1999). A natural starting point for any such extension would be to assume that purchase amount is independent of purchase timing (Schmittlein and Peterson 1994).

The BG/NBD easily lends itself to relevant generalizations, such as the inclusion of demographics or measures of marketing activity. However great care must be exercised when undertaking such extensions: to the extent that customer segments have been formed on the basis of past behavior (e.g., using the RFM framework) and these segments have been targeted with different marketing activities, we must be aware of econometric issues such as endogeneity bias (Shugan 2004) and sample selection bias. If such extensions are undertaken, the BG/NBD in its basic form would still serve as an appropriate (and hard-to-beat) benchmark model and should be viewed as the right starting point for any customer-base analysis exercise.

Appendix

In this appendix, we derive the expressions for $E[(X(t)]$ and $E(Y(t)\,|\,X = x, t_x, T)$. Central to these derivations is Euler's integral for the Gaussian hypergeometric function:

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

Derivation of $E[X(t)]$

To arrive at an expression for $E[X(t)]$ for a randomly-chosen customer, we need to take the expectation of (5) over the distribution of λ and p. First we take the expectation with respect to λ, giving us

$$ E(X(t)\,|\,r,\alpha,p) = \frac{1}{p} - \frac{\alpha^r}{p(\alpha+pt)^r} $$

The next step is to take the expectation of this over the distribution of p. We first evaluate

$$ \int_0^1 \frac{1}{p}\frac{p^{a-1}(1-p)^{b-1}}{B(a,b)}dp = \frac{a+b-1}{a-1} $$

Next, we evaluate

$$ \int_0^1 \frac{\alpha^r}{p(\alpha+pt)^r}\frac{p^{a-1}(1-p)^{b-1}}{B(a,b)}dp = \alpha^r \frac{1}{B(a,b)} \int_0^1 p^{a-2}(1-p)^{b-1}(\alpha+pt)^{-r}dp $$

letting $q = 1-p$ (which implies $dp = -dq$)

$$ = \left(\frac{\alpha}{\alpha+t}\right)^r \frac{1}{B(a,b)} \int_0^1 q^{b-1}(1-q)^{a-2}\big(1-\tfrac{t}{\alpha+t}q\big)^{-r}dq $$

which, recalling Euler's integral for the Gaussian hypergeometric function,

$$ = \left(\frac{\alpha}{\alpha+t}\right)^r \frac{B(a-1,b)}{B(a,b)}\,{}_2F_1\big(r,b;a+b-1;\tfrac{t}{\alpha+t}\big) $$

It follows that

$$E(X(t)\,|\,r,\alpha,a,b) = \frac{a+b-1}{a-1}\left[1-\left(\frac{\alpha}{\alpha+t}\right)^r {}_2F_1\big(r,b;a+b-1;\tfrac{t}{\alpha+t}\big)\right]$$

Derivation of $E(Y(t)\,|\,X = x, t_x, T)$

Let the random variable $Y(t)$ denote the number of purchases made in the period $(T, T+t]$. We are interested in computing the conditional expectation $E(Y(t)\,|\,X = x, t_x, T)$, the expected number of purchases in the period $(T, T+t]$ for a customer with purchase history $X = x, t_x, T$.

If the customer is active at T, it follows from (5) that

$$E(Y(t)\,|\,\lambda,p) = \frac{1}{p} - \frac{1}{p}e^{-\lambda p t} \tag{A1}$$

What is the probability that a customer is active at T? Given our assumption that all customers are active at the beginning of the initial observation period, a customer cannot drop out before he has made any transactions; therefore,

$$P(\text{active at } T\,|\,X = 0, T, \lambda, p) = 1$$

For the case where purchases were made in the period $(0, T]$, the probability that a customer with purchase history $(X = x, t_x, T)$ is still active at T, conditional on λ and p, is simply the probability that he did not drop out at t_x *and* made no purchase in $(t_x, T]$, divided by the probability of making no purchases in this same period. Recalling that this second probability is simply the probability that the customer became inactive at t_x, plus the probability he remained active but made no purchases in this interval, we have

$$P(\text{active at } T\,|\,X = x, t_x, T, \lambda, p) = \frac{(1-p)e^{-\lambda(T-t_x)}}{p + (1-p)e^{-\lambda(T-t_x)}}$$

Multiplying this by $[(1-p)^{x-1}\lambda^x e^{-\lambda t_x}]/[(1-p)^{x-1}\lambda^x e^{-\lambda t_x}]$ gives us

$$P(\text{active at } T\,|\,X = x, t_x, T, \lambda, p) = \frac{(1-p)^x \lambda^x e^{-\lambda T}}{L(\lambda, p\,|\,X = x, t_x, T)} \tag{A2}$$

where the expression for $L(\lambda, p \,|\, X = x, t_x, T)$ is given in (3). (Note that when $x = 0$, the expression given in (A2) equals 1.)

Multiplying (A1) and (A2) yields

$$\begin{aligned} E(Y(t) \,|\, X = x, t_x, T, \lambda, p) &= \frac{(1-p)^x \lambda^x e^{-\lambda T} \left(\frac{1}{p} - \frac{1}{p} e^{-\lambda p t} \right)}{L(\lambda, p \,|\, X = x, t_x, T)} \\ &= \frac{p^{-1}(1-p)^x \lambda^x e^{-\lambda T} - p^{-1}(1-p)^x \lambda^x e^{-\lambda(T+pt)}}{L(\lambda, p \,|\, X = x, t_x, T)} \end{aligned} \tag{A3}$$

(Note that this reduces to (A1) when $x = 0$, which follows from the result that a customer who made zero purchases in the time period $(0, T]$ must be assumed to be active at time T.)

As the transaction rate λ and dropout probability p are unobserved, we compute $E(Y(t) \,|\, X = x, t_x, T)$ for a randomly chosen customer by taking the expectation of (A3) over the distribution of λ and p, updated to take account of the information $X = x, t_x, T$:

$$\begin{aligned} &E(Y(t) \,|\, X = x, t_x, T, r, \alpha, a, b) = \\ &\qquad \int_0^1 \int_0^\infty E(Y(t) \,|\, X = x, t_x, T, \lambda, p) f(\lambda, p \,|\, r, \alpha, a, b, X = x, t_x, T) d\lambda \, dp \end{aligned} \tag{A4}$$

By Bayes theorem, the joint posterior distribution of λ and p is given by

$$f(\lambda, p \,|\, r, \alpha, a, b, X = x, t_x, T) = \frac{L(\lambda, p \,|\, X = x, t_x, T) f(\lambda \,|\, r, \alpha) f(p \,|\, a, b)}{L(r, \alpha, a, b \,|\, X = x, t_x, T)} \tag{A5}$$

Substituting (A3) and (A5) in (A4), we get

$$E(Y(t) \,|\, X = x, t_x, T, r, \alpha, a, b) = \frac{\mathsf{A} - \mathsf{B}}{L(r, \alpha, a, b \,|\, X = x, t_x, T)} \tag{A6}$$

where

$$\begin{aligned} \mathsf{A} &= \int_0^1 \int_0^\infty p^{-1}(1-p)^x \lambda^x e^{-\lambda T} f(\lambda \,|\, r, \alpha) f(p \,|\, a, b) d\lambda \, dp \\ &= \frac{B(a-1, b+x)}{B(a, b)} \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)(\alpha+T)^{r+x}} \end{aligned} \tag{A7}$$

and

$$
\begin{aligned}
\mathsf{B} &= \int_0^1 \int_0^\infty p^{-1}(1-p)^x \lambda^x e^{-\lambda(T+pt)} f(\lambda \,|\, r,\alpha) f(p \,|\, a,b) d\lambda \, dp \\
&= \int_0^1 \frac{p^{a-2}(1-p)^{b+x-1}}{B(a,b)} \left\{ \int_0^\infty \frac{\alpha^r \lambda^{r+x-1} e^{-\lambda(\alpha+T+pt)}}{\Gamma(r)} d\lambda \right\} dp \\
&= \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)B(a,b)} \int_0^1 p^{a-2}(1-p)^{b+x-1}(\alpha+T+pt)^{-(r+x)} dp
\end{aligned}
$$

letting $q = 1-p$ (which implies $dp = -dq$)

$$
= \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)B(a,b)(\alpha+T+t)^{r+x}} \int_0^1 q^{b+x-1}(1-q)^{a-2}\big(1 - \tfrac{t}{\alpha+T+t}q\big)^{-(r+x)} dq
$$

which, recalling Euler's integral for the Gaussian hypergeometric function,

$$
= \frac{B(a-1,b+x)}{B(a,b)} \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)(\alpha+T+t)^{r+x}} {}_2F_1\big(r+x, b+x; a+b+x-1; \tfrac{t}{\alpha+T+t}\big) \tag{A8}
$$

Substituting (6), (A7) and (A8) in (A6) and simplifying, we get

$$
E(Y(t) \,|\, X = x, t_x, T, r, \alpha, a, b) =
$$

$$
\frac{\dfrac{a+b+x-1}{a-1}\left[1 - \left(\dfrac{\alpha+T}{\alpha+T+t}\right)^{r+x} {}_2F_1\big(r+x, b+x; a+b+x-1; \tfrac{t}{\alpha+T+t}\big)\right]}{1 + \delta_{x>0}\dfrac{a}{b+x-1}\left(\dfrac{\alpha+T}{\alpha+t_x}\right)^{r+x}}
$$

References

Albers, Sönke (2000), "Impact of Types of Functional Relationships, Decisions, and Solutions on the Applicability of Marketing Models," *International Journal of Research in Marketing*, **17** (2–3), 169–175.

Balasubramanian, S., S. Gupta, W. Kamakura, and M. Wedel (1998), "Modeling Large Datasets in Marketing," *Statistica Neerlandica*, **52** (3), 303–323.

Blattberg, Robert C., Gary Getz, and Jacquelyn S. Thomas (2001), *Customer Equity*, Boston, MA: Harvard Business School Press.

Colombo, Richard and Weina Jiang (1999), "A Stochastic RFM Model," *Journal of Interactive Marketing*, **13** (Summer), 2–12.

Fader, Peter S. and Bruce G. S. Hardie, (2001), "Forecasting Repeat Sales at CDNOW: A Case Study," *Interfaces*, **31** (May–June), Part 2 of 2, S94–S107.

Jain, Dipak and Siddhartha S. Singh (2002), "Customer Lifetime Value Research in Marketing: A Review and Future Directions," *Journal of Interactive Marketing*, **16** (Spring), 34–46.

Lozier D. W. and F. W. J. Olver (1995), "Numerical Evaluation of Special Functions," in Walter Gautschi (ed.), *Mathematics of Computation 1943–1993: A Half-Century of Computational Mathematics*, Proceedings of Symposia in Applied Mathematics, Providence, RI: American Mathematical Society.

Mulhern, Francis J. (1999), "Customer Profitability Analysis: Measurement, Concentration, and Research Directions," *Journal of Interactive Marketing*, **13** (Winter), 25–40.

Niraj, Rakesh, Mahendra Gupta, and Chakravarthi Narasimhan (2001), "Customer Profitability in a Supply Chain," *Journal of Marketing*, **65** (July), 1–16.

Reinartz, Werner and V. Kumar (2000), "On the Profitability of Long-Life Customers in a Non-contractual Setting: An Empirical Investigation and Implications for Marketing," *Journal of Marketing*, **64** (October), 17–35. Reinartz, Werner and V. Kumar (2003), "The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration," *Journal of Marketing* **67** (January), 77–99.

Schmittlein, David C., Donald G. Morrison, and Richard Colombo (1987), "Counting Your Customers: Who They Are and What Will They Do Next?" *Management Science*, **33** (January), 1–24.

Schmittlein, David C. and Robert A. Peterson (1994), "Customer Base Analysis: An Industrial Purchase Process Application," *Marketing Science*, **13** (Winter), 41–67.

Shugan, Steven M. (2004), "Endogeneity in Marketing Decision Models," *Marketing Science*, **23** (Winter), 1–3.

Wu, Couchen and Hsiu-Li Chen (2000), "Counting Your Customers: Compounding Customer's In-store Decisions, Interpurchase Time, and Repurchasing Behavior," *European Journal of Operational Research*, **127** (1) 109–119.