Counting Your Customers: Who Are They and What Will They Do Next?
Author(s): David C. Schmittlein, Donald G. Morrison and Richard Colombo
Source: *Management Science*, Vol. 33, No. 1 (Jan., 1987), pp. 1-24
Published by: INFORMS
Stable URL: http://www.jstor.org/stable/2631608

MANAGEMENT SCIENCE
Vol. 33, No. 1, January 1987
Printed in U.S.A.

COUNTING YOUR CUSTOMERS: WHO ARE THEY AND WHAT WILL THEY DO NEXT?

DAVID C. SCHMITTLEIN, DONALD G. MORRISON AND RICHARD COLOMBO
The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104
Graduate School of Business, Columbia University, New York, New York 10027
Department of Marketing, Tisch Hall, New York University, New York, New York 10002

This article is concerned with counting and identifying those customers who are still active. The issue is important in at least three settings: monitoring the size and growth rate of a firm's ongoing customer base, evaluating a new product's success based on the pattern of trial and repeat purchases, and targeting a subgroup of customers for advertising and promotions. We develop a model based on the number and timing of the customers' previous transactions. This approach allows computation of the probability that any particular customer is still active. Several numerical examples are used to illustrate applications of the model.

(MARKETING; CONSUMER BEHAVIOR; POISSON PROCESS; PROBABILITY MIXTURE MODELS; NEW PRODUCT INTRODUCTIONS; MARKET SEGMENTATION; BROKERAGE FIRMS)

1. Introduction

Suppose that you have recently been put in charge of the retail brokerage side of a large firm such as Merrill Lynch. The financial services industry has been in a continual state of flux for the last decade. Serious marketing research is relatively new to this industry. You have a list of customers who have ever done business with the firm, as well as information on the frequency and timing of each customer's transactions. Therefore, as a starting point you ask yourself four very simple questions:

1. How many retail customers does the firm now have?
2. How has this customer base grown over the past year?
3. Which individuals on this list most likely represent active customers? Inactive customers?
4. What level of transactions should be expected next year by those on the list, both individually and collectively?

Of course these questions are not unique to the brokerage industry. All sorts of organizations have lists that include some former "customers" who are no longer "active", e.g.

— catalog mailing lists,
— church directories,
— dentists' and beauty salons' files,
— department store charge card records, and
— triers of a new grocery product (members of a diary or UPC scanner panel).

All executives who ask customer base questions are implicitly asking questions about "active" customers, i.e., individuals who have not become disenchanted, taking their business or patronage elsewhere. Determining the probability that a customer with a given pattern of transactions is still active is the main focus of this research, and the model developed to do so will provide answers to all four questions above.

The total list of customers who have ever done business with you is a good starting point. If we could attach a probability to each customer, i.e., the probability that this

* Accepted by John R. Hauser; received May 20, 1985. This paper has been with the authors 5 months for 1 revision.

0025-1909/87/3301/0001$01.25

particular customer is still active, questions 1 and 3 above would be answered immediately. At the current point in time we would merely add up all of the individual customer probabilities and this sum would be our estimate of the current number of active customers. For example, if we had three customers with probabilities of being active of

$$\begin{array}{rr} & 0.9 \\ & 0.7 \\ & \underline{0.4} \\ \text{Total} & 2.0 \end{array}$$

we would estimate that we had two active customers.

We would use the same procedure with data as of one year ago in order to estimate our active customer base back then. This will provide the benchmark for estimating the growth rate over the past year (question 2). The role of this probability is making predictions of future transactions (question 4) is more subtle and we leave the details to §2. In principle, though, the idea is simple. We have one expectation (which will be based on past purchase levels) if the customer is still active, and a different expectation (equal to zero) if the customer is inactive. So the probability that the customer is active allows us to weight these two predictions appropriately.

The counting of active customers is difficult for one—and only one—reason. The vast majority of customers who leave the firm do not notify the firm when they leave. A client may become disenchanted with the investment advice and move to another firm. The client may tell the account executive, but is unlikely to notify the firm. Even if the customer leaves and physically takes the stock from the firm's custody, this information may not get entered into the master file of accounts. Customers can also leave by following their account executive to another firm, by moving to another city and, in the ultimate form of leaving, they can die.

Assuming that the firm has not been notified when customers leave, the firm will have two kinds of transaction information on each customer:

i. The activity (number of transactions, commission dollars, etc.) during the past time period being analyzed.

ii. The time between the last transaction and the end of the observation period.

Consider two different customers, each having made four transactions during the past calendar year. Customer A made all four purchases in January and no purchases during the February through December period. Customer B made one purchase in each of the following months: March, June, September and December. Clearly B is very likely to be still an active customer while it is quite likely that A is either out of the market or doing business at another firm. The goal of this article is to quantify the above qualitative assessments. We will continue to motivate our efforts within the broker-client framework. The examples listed at the outset of this paper indicate, however, that the type of probabilistic modelling needed to count active customers has applicability far beyond the brokerage industry.

The remainder of the paper is organized as follows. §2 will develop and justify the models for the purchasing process and the duration time that a customer remains active. It will also contain the major contribution, namely the derivation of the conditional probability:

$$P(\text{Customer Active}|\text{Purchasing Information}).$$

The key conceptual equations will be presented in the text with the somewhat tedious steps of the derivations relegated to the Appendix. The application of these results to the questions posed at the outset is the subject of §3. §4 will have some numerical examples that will challenge and perhaps contradict the reader's intuition. §5 is concerned with estimating the model's parameters, and §6 will discuss other fruitful areas for applying our model, along with implementation issues that will surely arise.

2. Model Development: Computing the Probability That a Customer Is Alive

Our model will consider only purchasing or transaction events. That is, the amount purchased or commissions generated would be analyzed separately, or added as an extension to our approach. Each such transaction event will be termed a "purchase." A customer who is still active will be termed "Alive," while customers who have left for whatever reason will be termed "Dead." A customer who is alive at time 0 is observed up to time T. During this observation the customer will have made X purchases with the last purchase coming at t, $0 < t \leqq T$. Hence, the information on this customer contains 3 elements:

$$\text{Information} = (X, t, T).$$

For the model developed in this section, we show in the Appendix that these 3 elements represent all of the relevant information from the customer's transaction pattern. Our goal is to calculate the conditional probability

$$P(\text{Alive}|\text{Information}).$$

Before doing so we must make some assumptions about the purchase event process and the time that the customer stays active (alive). These five assumptions are given in words here. We will attempt to convince the reader that these assumptions are sensible, and then present the formalized mathematics of the model.

Individual Customer

1. *Poisson Purchases*: While alive, each customer makes purchases according to a Poisson process with rate λ.
2. *Exponential Lifetime*: Each customer remains alive for a lifetime which has an exponentially distributed duration with death rate μ.

Heterogeneity Across Customers

3. *Individuals' Purchasing Rates Distributed Gamma*: The purchasing rate λ for the different customers is distributed according to a gamma distribution across the population of customers.
4. *Death Rates Distributed Gamma*: The customer's death rates μ are distributed according to a different gamma distribution across customers.
5. *Rates λ and μ Are Independent*: The purchasing rates λ and the death rates μ are distributed independently of each other.

Justification of the Assumptions

The Poisson purchasing process implies exponentially distributed interpurchasing times with the resulting lack of memory property. For brokerage clients this seems like a reasonable first approximation. That is, each day there is a constant probability of being stimulated to make a transaction. In addition, one of us (DGM) has examined transactions data at Merrill Lynch and the Poisson assumption at the individual level appears to be a good approximation. For frequently purchased consumer goods the Poisson assumption has a long validated history (Ehrenberg 1972). The Poisson assumption would not be as good with catalog purchases since the new editions of the catalog may be mailed out on a very regular schedule.

The exponential duration for staying "alive," i.e., active or interested, also seems entirely reasonable. The events that could trigger "death" (a move, a financial setback, a lifestyle change, etc.) may arrive in a Poisson manner. Even if the specific death events (e.g., a move to another town) are not Poisson, the arrival of all possible death events is the superposition of the individual events, and this superposition process is driven closer to a Poisson process (Karlin and Taylor 1975, p. 221).

Some sort of heterogeneity assumptions are mandated. There are frequent purchasers and infrequent purchasers. Some people will have more and faster arriving potential death events than will other customers. The gamma is a flexible distribution and can capture the spirit of most of the reasonable distributions on the purchasing rate λ and the death rate μ.

Finally, there is no a priori reason to favor a positive correlation between λ and μ over a negative correlation. That is, a heavy purchaser may have more frequent opportunities to be disenchanted by the product (e.g., bad investment advice) and "die." On the other hand, this heavy purchaser is probably more strongly attached to the product (e.g., the stock market) and hence less easily disenchanted. Independence between λ and μ seems like a good assumption, subject to future tests.

A mathematical statement of our assumptions and the derivation of P(Alive|Information) follows. At this point we hope that the reader is convinced that deriving this conditional probability helps solve important problems and that our underlying model is not only mathematically tractable but also realistic.

Model Specification

In developing the model and the quantities of interest it will be easiest to begin at the individual level—that is, with a single customer having a given purchase rate λ and death rate μ. As above, X is the number of purchases the customer makes in time period $(0, T]$ with the last purchase coming at time $t \leqq T$. It is assumed that the customer is "alive" (active) at time 0. This will generally be satisfied if we take time = 0 as the point at which the customer makes his or her initial purchase. Notice, then, that the relation between the period $(0, T]$ and calendar time will vary from customer to customer, depending on when this person's initial purchase occurred.

Let the (unobserved) time at which the customer becomes inactive be denoted τ. For any time $T > 0$, if the customer is still alive at T (so $\tau > T$) the number of purchases in $(0, T]$ has the Poisson distribution:

$$P[X = x|\lambda, \tau > T] = e^{-\lambda T}\frac{(\lambda T)^x}{x!}; \qquad x = 0, 1, 2, \ldots, \tag{1}$$

with expected value and variance

$$E[X|\lambda, \tau > T] = \lambda T; \qquad \text{Var}\,[X|\lambda, \tau > T] = \lambda T.$$

The time τ until becoming inactive is assumed to follow the exponential distribution, with p.d.f.:

$$f(\tau|\mu) = \mu e^{-\mu\tau}; \qquad \tau > 0; \qquad \text{where}$$

$$E[\tau|\mu] = 1/\mu; \qquad \text{Var}\,[\tau|\mu] = 1/\mu^2. \tag{2}$$

For this particular individual with purchase rate λ, death rate μ, $X = x$ transactions in $(0, T]$, and the last purchase at t, the desired probability P[Customer Alive|Purchasing Information] is derived in the Appendix. From equation (A10) the formula is

$$P[\tau > T|\lambda, \mu, X = x, t, T] = \frac{1}{1 + (\mu/(\lambda + \mu))[e^{(\lambda+\mu)(T-t)} - 1]}, \tag{3}$$

where the event $\tau > T$ indicates that the customer is still alive at T. In the situation where no transactions were made in $(0, T]$ (i.e. $X = 0$) this probability $P[\tau > T|\lambda, \mu, X = 0, —, T]$ is obtained by substituting $t = 0$ in (3).

Unfortunately, equation (3) is not very helpful. The parameters λ and μ which determine this quantity are not known, nor is there generally enough past information on this particular customer to estimate λ and μ for this individual alone. Since other customers have different purchase rates and death rates their purchase patterns cannot simply be pooled to estimate λ and μ.

Instead, we wish to derive P[Alive|Information] for an individual chosen at random from all of the firm's customers, whose (λ, μ)-values are unknown. While there is not enough information to reliably estimate these 2 parameters for each person, there will generally be enough to estimate the distribution of (λ, μ) over customers. That estimate will be used as a prior distribution and combined with (3) to yield the desired probability. This approach, estimating a prior distribution from the available data, is usually called an empirical Bayes method (Maritz 1970, Morris 1983).

The purchase rates and death rates are assumed to follow different gamma distributions over customers, with p.d.f.'s

$$g(\lambda|r, \alpha) = \frac{\alpha^r}{\Gamma(r)} \lambda^{r-1} e^{-\alpha\lambda}; \qquad \lambda > 0; \qquad r, \alpha > 0;$$

$$E[\lambda|r, \alpha] = r/\alpha; \qquad \text{Var}\,[\lambda|r,\alpha] = r/\alpha^2; \qquad \text{and} \tag{4}$$

$$h(\mu|s, \beta) = \frac{\beta^s}{\Gamma(s)} \mu^{s-1} e^{-\beta\mu}; \qquad \mu > 0; \qquad s, \beta > 0;$$

$$E[\mu|s, \beta] = s/\beta; \qquad \text{Var}\,[\mu|s, \beta] = s/\beta^2. \tag{5}$$

The coefficient of variation (i.e. standard deviation divided by the mean) of λ is $r^{-1/2}$. r is thus an index of homogeneity in purchase rates over the population of customers. In the same way, s indicates the homogeneity in death rates.

Together with the assumption that λ and μ are independent, equations (1), (2), (4) and (5) fully specify the model considered here. Statements (1) and (4) imply that purchases made while a customer is active follow the NBD model (Ehrenberg 1972) and have the distribution

$$P\,[X = x|r, \alpha, \tau > T] = \binom{x+r-1}{x}\left(\frac{\alpha}{\alpha+T}\right)^r\left(\frac{T}{\alpha+T}\right)^x; \quad x = 0, 1, 2, \ldots, \quad \text{where} \tag{6}$$

$$E[X|r, \alpha, \tau > T] = \frac{rT}{\alpha}; \qquad \text{and} \tag{7}$$

$$\text{Var}\,[X|r, \alpha, \tau > T] = \frac{rT}{\alpha} + \frac{rT^2}{\alpha^2}; \tag{8}$$

for the number of purchases X made in time period $[0, T]$. Also, as a consequence of (2) and (5) "deaths" for a sample of customers follow the Pareto distribution of the second kind (Johnson and Kotz 1970, p. 233), with p.d.f.

$$f(\tau|s, \beta) = \frac{s}{\beta}(\beta/(\beta+\tau))^{s+1}, \qquad \tau > 0;$$

$$E[\tau|s, \beta] = \frac{\beta}{s-1}, \qquad s > 1. \tag{9}$$

The expected value of τ diverges to infinity if $s \leqq 1$. This combined purchase event/duration model will be called the Pareto/NBD. It has four parameters: r, α, s and β.

P[Alive|Information]

For a randomly chosen individual having purchase pattern ($X = x, t, T$) the probability that this customer is still active is

$$P[\tau > T | r, \alpha, s, \beta, X = x, t, T]$$
$$= \int_0^\infty \int_0^\infty P[\tau > T | \lambda, \mu, X = x, t, T] f(\lambda, \mu | r, \alpha\, s, \beta, X = x, t, T) d\lambda d\mu. \quad (10)$$

Expression (10) is the weighted average over λ and μ of the individual-level probabilities $P[\tau > T | \lambda, \mu, X = x, t, T]$ given in (3). The weights are the likelihood that the particular (λ, μ) value holds for that customer. That is, $f(\lambda, \mu | r, \alpha, s, \beta, X = x, t, T)$ is the updated distribution of λ and μ given the observed purchase pattern.

In the Appendix we derive this updated distribution, and use (3) and (10) to compute the expression for the desired quantity P[Customer Active|$X = x, t, T$]. This probability that the customer is still alive is listed next in equations (11), (12) and (13). The result varies depending on whether $\alpha > \beta$, $\alpha < \beta$ or $\alpha = \beta$:

Case 1: $\alpha > \beta$.

$$P[\tau > T | r, s, \alpha > \beta, X = x, t, T]$$
$$= \left\{1 + \frac{s}{r+x+s}\left[\left(\frac{\alpha+T}{\alpha+t}\right)^{r+x}\left(\frac{\beta+T}{\alpha+t}\right)^{s} F(a_1, b_1; c_1; z_1(t)) - \left(\frac{\beta+T}{\alpha+T}\right)^{s} F(a_1, b_1; c_1; z_1(T))\right]\right\}^{-1} \quad \text{where} \quad (11)$$

$$a_1 = r + x + s; \qquad b_1 = s + 1;$$
$$c_1 = r + x + s + 1; \qquad z_1(y) = \frac{\alpha - \beta}{\alpha + y}.$$

Case 2: $\alpha < \beta$.

$$P[\tau > T | r, s, \alpha < \beta, X = x, t, T]$$
$$= \left\{1 + \frac{s}{r+x+s}\left[\left(\frac{\alpha+T}{\beta+t}\right)^{r+x}\left(\frac{\beta+T}{\beta+t}\right)^{s} F(a_2, b_2; c_2; z_2(t)) - \left(\frac{\alpha+T}{\beta+T}\right)^{r+x} F(a_2, b_2; c_2; z_2(T))\right]\right\}^{-1} \quad \text{where} \quad (12)$$

$$a_2 = r + x + s; \qquad b_2 = r + x;$$
$$c_2 = r + x + s + 1; \qquad z_2(y) = \frac{\beta - \alpha}{\beta + y}.$$

Case 3: $\alpha = \beta$.

$$P[\tau > T | r, s, \alpha = \beta, X = x, t, T] = \left\{1 + \frac{s}{r+x+s}\left[\left(\frac{\alpha+T}{\alpha+t}\right)^{r+x+s} - 1\right]\right\}^{-1}. \quad (13)$$

In (11) and (12) $F(a, b; c; z)$ is the Gauss hypergeometric function (Abramowitz and Stegun 1972, p. 558). It can be computed using either numerical integration or the algorithms in Luke (1977). For the examples discussed subsequently in §4, $F(a, b; c; z)$ was calculated via numerical integration. As noted in the Appendix, when no transactions are observed in the period $(0, T]$ the probability that the customer is still active is obtained by substituting $x = 0$ and $t = 0$ in (11), (12) or (13).

Additional Model Properties

Although the probability given in (11)–(13) is the main result of interest, a few other properties of the Pareto/NBD model will be helpful. One is the expected number of transactions by a customer in a time period $(0, T]$. Unlike the case for the NBD model in (6), this customer may have become inactive at some point in the time interval. These expectations will turn out to be useful for predicting future purchase levels (question #4 in the Introduction) and also in estimating the four model parameters (r, α, s, β).

The expected number of purchases for a randomly chosen customer, $E[X|r, \alpha, s, \beta, T]$ is easily derived by conditioning on the amount of time θ that the customer has to make those purchases in the period $(0, T]$. So θ is τ if the customer becomes inactive before time T, and is T if he/she doesn't; i.e., the minimum of τ and T. Using (7) the expected value of X given θ is

$$E[X|r, \alpha, s, \beta, T, \theta] = \frac{r}{\alpha}\theta. \tag{14}$$

Taking the expectation of (14) over the possible θ-values yields

$$E[X|r, \alpha, s, \beta, T] = \frac{r}{\alpha}E[\theta|r, \alpha, s, \beta, T]. \tag{15}$$

The distribution of θ depends only on the duration time parameters s and β. The expected value of θ can be derived easily using (9) and the relation described above between θ and τ:

$$E[\theta|r, \alpha, s, \beta, T] = \frac{\beta}{s-1}\left[1 - \left(\frac{\beta}{\beta+T}\right)^{s-1}\right]. \tag{16}$$

Substituting (16) in (15) the expected number of purchases is

$$E[X|r, \alpha, s, \beta, T] = \frac{r\beta}{\alpha(s-1)}\left[1 - \left(\frac{\beta}{\beta+T}\right)^{s-1}\right]. \tag{17}$$

Notice that to compute the expected number of purchases for any length of time T one need know only the three quantities r/α, s and β.

It will also be useful to have the expression for the variance of the number of transactions made in a given interval $(0, T]$. As for the expectation above it is convenient to condition on the amount of time available θ. Using a general relation for conditional means and variances (Feller 1971, p. 167) the desired variance is

$$\mathrm{Var}\,[X|r, \alpha, s, \beta, T] = E_\theta[\mathrm{Var}\,[X|r, \alpha, s, \beta, T, \theta]] + \mathrm{Var}_\theta\,[E[X|r, \alpha, s, \beta, T, \theta]] \tag{18}$$

where E_θ and Var_θ denote the expectation and variance over θ. Using (8) and (14), (18) becomes

$$\begin{aligned}\mathrm{Var}\,[X|r, \alpha, s, \beta, T] &= E_\theta\left[\frac{r\theta}{\alpha} + \frac{r\theta^2}{\alpha^2}\middle| r, \alpha, s, \beta, T\right] + \frac{r^2}{\alpha^2}\mathrm{Var}_\theta\,[\theta|r, \alpha, s, \beta, T] \\ &= \frac{r}{\alpha}E[\theta|s, \beta, T] - \frac{r^2}{\alpha^2}(E[\theta|s, \beta, T])^2 + \frac{r(r+1)}{\alpha^2}E[\theta^2|s, \beta, T]. \end{aligned} \tag{19}$$

So the variance can be computed by substituting (16) and the following result in (19):

$$E[\theta^2|s, \beta, T] = \frac{2\beta}{s-1}\left[\frac{\beta}{s-2} - \frac{\beta}{s-2}\left(\frac{\beta}{\beta+T}\right)^{s-2} - T\left(\frac{\beta}{\beta+T}\right)^{s-1}\right]. \tag{20}$$

As for expression (16), (20) is derived using (9) and the fact that θ is the minimum of τ and T.

Equations (17) and (19) give the first two moments for the distribution of the observed number of purchases X in time period $(0, T]$. In addition, the probability distribution itself can be derived and is given in the Appendix (A40, A43 and A45).

Finally, it will be helpful to consider the conditional distribution for the number of purchases X^* in a future time period of duration T^*, given the information (X, t, T). That is, we know that a person has made X purchases in time period $(0, T]$ with the last transaction at time t. Conditioned on this information (X, t, T) what is the distribution for the number of purchases in period $(T, T + T^*]$? This conditional distribution will allow us to answer the fourth question posed in the Introduction (i.e. "What level of transactions should be expected next year by those on the customer list, both individually and collectively?")

Of course, in the Pareto/NBD model if a customer is no longer active at time T then the number of purchases made in any future period is going to be zero. On the other hand, if the customer is still active at T the conditional distribution of X^* is very easy to obtain. For an individual, given no "death" in $(0, T]$ the time from T until becoming inactive still has the exponential distribution with death rate μ, due to the exponential's lack of memory property. Also, the updated distribution of μ across individuals is still gamma, but with the new parameters $s^* = s$, $\beta^* = \beta + T$ (Morrison 1978). Since the death rate and purchase rate are assumed to be independent, the information on purchases made in $(0, T]$ is irrelevant for updating the death process, given that the individual is still alive at T. So a randomly chosen individual who is still active at T has a death process proceeding from T that again follows a Pareto distribution, but with updated parameters $s^* = s$ and $\beta^* = \beta + T$.

To complement this updated death process, the updated purchase process for the customer who is still alive at T is also easy to derive. Since there was no death in $(0, T]$ the purchases made in that period follow the NBD model (6) for a randomly chosen individual. Consequently, proceeding from time T, the purchases made by an individual with purchase rate λ will still follow a Poisson process. The updated distribution of λ across individuals, given that an NBD process for $(0, T]$ resulted in $X = x$ purchases, is again gamma with updated parameters $r^* = r + x$, $\alpha^* = \alpha + T$ (Morrison 1968). So, while alive, the purchases made by this randomly chosen individual in $(T, T + T^*]$ will again follow the NBD model, with new parameters $r^* = r + x$, $\alpha^* = \alpha + T$.

Summarizing, a person "dead" at T will never make additional purchases. A person "alive" at T will again follow the Pareto/NBD process during the period $(T, T + T^*]$, but with new parameters $r^* = r + x$, $\alpha^* = \alpha + T$, $s^* = s$, $\beta^* = \beta + T$. For this person it is as if the process started all over at time T with parameters $(r^*, \alpha^*, s^*, \beta^*)$. So the desired distribution of future purchases X^* in period $(T, T + T^*]$ conditioned only on the observed information (X, t, T) is

$$\begin{aligned} P[X^* = x^* | r, \alpha, s, \beta, X, t, T, T^*] &= P[X^* = x^* | r, \alpha, s, \beta, X, t, T, T^*, \text{“Alive” at } T] \\ &\quad \times P[\text{“Alive” at } T | r, \alpha, s, \beta, X, t, T, T^*] \\ &= P[X^* = x^* | r + x, \alpha + T, s, \beta + T, T^*] P[\tau > T | r, \alpha, s, \beta, X, t, T]. \quad (21) \end{aligned}$$

The first term on the right is the updated Pareto/NBD purchase probability which is computed by substituting these new parameters in the original expression for the probability distribution (A40), (A43) or (A45). The second term in (21) is the probability that the customer is still alive at T given the information (X, t, T). The formulas for this probability were given in (11)–(13) above.

Using (21) the expected number of purchases in the period $(T, T + T^*]$ for a customer with purchase profile $(X = x, t, T)$ can also be derived. This expectation is

$$E[X^*|r, \alpha, s, \beta, X = x, t, T, T^*] = E[X^*|r + x, \alpha + T, s, \beta + T, T^*]$$
$$\times P[\tau > T|r, \alpha, s, \beta, X = x, t, T]. \quad (22)$$

The first term on the right is the usual expectation for Pareto/NBD purchases (with updated parameters) given in (17), and the second term is again the probability of being alive at T, listed in (11)–(13). If we were to add up the expression (22) for all individuals in our customer list we would have the expected total number of purchases in the period $(T, T + T^*]$ to be made by this group.

Summary

In addition to describing the purchase/death model and justifying its assumptions, we have developed the following results in this section:

• P[Customer Active|Purchasing Information]. (Equations 11–13). This probability is the main result of this paper. As shown in the next section it can be used to answer the first three managerial questions raised at the outset of this article.

• Expectation and variance for the number of transactions made by a customer in $(0, T]$. (Equations 17 and 19, respectively). These two moments are useful process descriptors and will play a role in parameter estimation in section 5.

• Probability distribution for the number of transactions in $(0, T]$. (Equations A40, A43 and A45). Another useful descriptor of the purchase process and a possible basis for maximum likelihood parameter estimates.

• The distribution and expectation for the future number of transactions, given an observed purchase pattern (X, t, T). (Equations 21 and 22, respectively.) This is the key quantity for answering the fourth managerial question raised in the Introduction.

The next section discusses in more detail some managerial applications for the results here.

3. Using the Pareto/NBD Model to Solve Management Problems

We return now to the managerial issues that motivated this modelling effort. The four questions raised in the introduction concerned the size of the firm's customer base, the rate of growth or decline in that base over time, the identification of active vs. inactive individuals, and expectations about future transaction levels. This section describes some situations where these questions are important. It also details the role of the Pareto/NBD model in their solution.

Throughout the discussion it is assumed that

i. this model reasonably represents the purchase/death process, and

ii. the analyst has values for the four parameters (r, α, s, β) that specify the Pareto/NBD.

The general appropriateness of the Pareto/NBD was addressed in §2 and will also be indicated by the numerical examples in §4 and the discussion in §6. Methods for obtaining estimates of (r, α, s, β) will be the subject of §5.

The Number of Active Customers

As indicated in the introduction the expected number of active customers is just the sum over all individuals of P[Customer Active|Purchasing Information]. At one point in (calendar) time, say December 31, 1985, the amount of time T for which a customer is observed will vary from customer to customer. Some may have made their first transaction in November 1985, while others have been customers for several years. Also varying across individuals will be the number of observed purchases X and the elapsed time t between the beginning of that person's observation period and the last observed purchase.

Let subscript i refer to the information for the ith customer and recall that τ is the (unobserved) time until a customer becomes inactive. Then the expected number of active customers at (calendar) time C is

$$E[\# \text{ active at } C] = \sum_{i=1}^{M} P[\tau > T_i | r, \alpha, s, \beta, X_i = x_i, t_i, T_i] \tag{23}$$

where M is the number of customers observed, customer i's initial purchase came at (calendar) time $C - T_i$, and the quantities being summed are the probabilities from (11)–(13) that each customer is still active at C.

Of particular interest is the rate of change in the number of active customers over calendar time, which is easily computed using (23) for varying C-values. This rate of growth or decline is an important diagnostic regarding the basic health of a business, and since "deaths" are unobserved it can only be calculated via a model such as the Pareto/NBD. We should stress that simply measuring the growth rate of the "apparent customer base," i.e. the set of all customers with whom the firm has ever done business, is not an acceptable substitute. Such an index records only the number of new clients attracted to the firm, and ignores the length of time that each remains.

Selecting Active Customers

The expected number of active customers (23) is an aggregate property of the customer base, since it is not concerned with deciding the activity or inactivity of any particular customer. In some situations, however, the ability to make this decision about each individual is important. For example a firm may wish to pare its mailing list of the inactive customers in order to reduce database management costs and mailing costs.

In this situation the Pareto/NBD's probability P[Customer Active|Purchasing Information] in (11)–(13) can be used to rank order the customers. This is sufficient if the total number of customers to retain has been predetermined. Of course, if the costs and benefits of the four possible decision outcomes (Retain/Active, Retain/Inactive, Delete/Active and Delete/Inactive) can be specified then the model can be used to decide the optimal number of customers to use as well.

This highlights a major distinction between model-based approaches for identifying active customers (like the Pareto/NBD) and heuristic indices of activity that are sometimes used. A heuristic is hard to apply in a decision theoretic framework as it does not generate a probability of being active. Difficulties also arise in validating a heuristic for an unobservable event (becoming inactive) while both the assumptions and predictive effectiveness of the Pareto/NBD are open to scrutiny and, as a result, to improvement.

Furthermore, it is not easy to suggest a reasonable heuristic that uses all of the available and relevant information (X, t, T). One common approach calls a customer inactive if at least $t_{\max}$ units of time have elapsed since the last transaction (i.e. $T - t > t_{\max}$). So it ignores the number of purchases made X and the total time available T. Within the Pareto/NBD assumptions it can be shown that this simple heuristic is never appropriate for selecting active customers unless purchase rates are identical across all individuals (so $r \to \infty$) and death rates are also the same for all customers (so $s \to \infty$).

Predicting Future Transaction Levels

Another advantage of model-based approaches over heuristic indices of activity is the ability with the former to predict future purchasing behavior. As indicated in the last section the expected number of purchases in period $(T, T + T^*]$ for an individual with observed behavior (X, t, T) is given by equation (22). The expected number of transactions in this period that will be generated by the current customer base is obtained by summing (22) over all customers.

These results are useful in a wide variety of situations. For an ongoing business such as the brokerage firm discussed in the Introduction these expectations will indicate the number of new customers needed over time to attain a target sales or growth rate. They can also be used to select a subset of customers for special attention (e.g. to receive a promotion or be dropped from a mailing list as above). That is, it can be more appropriate to select customers based on expected future purchasing rather than on whether they are still "alive" (since a customer may still be active but have a very low purchase rate).

Finally, the Pareto/NBD can be helpful in analyzing the introduction of a consumer nondurable product. The model's expectations of future purchases, when combined with a model for initial trial, will predict the success or failure of the introduction. Equally important, the model's parameter values (r, α, s, β) can identify the reasons for this success or failure. For example, are disappointing sales levels due to a low average purchase rate (r/α) or a high average dropout rate (s/β)? The segmentability of the market based on purchase rates and death rates is also indicated by r and s, respectively. By allowing individual differences in both purchase rates and death rates the Pareto/NBD will capture the empirical regularities in new product repeat purchases noted by Eskin (1973), Eskin and Malec (1976) and Kalwani and Silk (1980). We next present some numerical illustrations of the Pareto/NBD model's characteristics.

4. Numerical Examples

The probability P[Customer Active|Purchasing Information] given in (11)–(13) is the most important quantity derived in this paper. It is used to calculate the number of active customers, identify those who are active, and (through (21) and (22)) to predict future transaction levels. This probability is listed in Table 1 for customers with a variety of purchase histories. We will be interested in the general pattern of these Pareto/NBD results. They confirm our intuition about the kind of customer that should be called "active," while resolving some cases where our intuition may not be so effective.

The four Pareto/NBD parameters (r, α, s, β) that characterize the purchase/death process for these customers are also given in the table. Based on these parameters, the average purchase rate λ while customers are active is $r/\alpha = 1$ unit per time period. The relatively small r-value of 0.415 indicates substantial differences in purchase rates across customers. This value was chosen to be consistent with previous empirical estimates of r for consumer packaged goods (Morrison and Schmittlein 1981).

The average death rate μ is $s/\beta = 0.5$. Recall that the time until a customer becomes inactive is exponentially distributed with rate μ, and expectation $1/\mu$. Then an "average" customer having $\mu = 0.5$ would be expected to remain active for 2 units of time. The relatively small value of 0.3 for s says that the death rate also varies greatly from customer to customer. With these (s, β) values it is expected that half of the customers will have become inactive after $(2^{1/s} - 1)\beta = 5.4$ units of time (Schmittlein and Morrison 1981).

Besides the parameters (r, α, s, β), the probability that a customer is active depends on the three pieces of information (X, t, T). Each column in Table 1 refers to a particular amount of elapsed time $T - t$ since the last observed transaction. The rows indicate the time t of the last observed purchase. The number of purchases X made in $(0, T]$, for $X = 0, 1, 2, 3, 4, 5$, is indicated by the six entries in each cell of the table. So, for example, the probability is 0.347 that a customer is still active at time $T = 1.25$ if 3 purchases were made in $(0, 1.25]$ with the last transaction occurring at $t = 0.25$. On the other hand, someone making the same number of purchases with the last occurring at $t = 1$ has a 0.944 probability of being alive at $T = 1.25$. The first entry in each cell, corre-

TABLE 1

P[Customer Active|Purchasing Information] at Time T

Information: X transactions in $(0,T]$ with the last transaction at t.

NBD Parameters: $r = 0.415$, $\alpha = 0.415$
Pareto Parameters: $s = 0.300$, $\beta = 0.600$

Cell Entries: $x = 0$, $x = 1$, $x = 2$, $x = 3$, $x = 4$, $x = 5$

Time t of Most Recent Transaction	Cell entry	Time $T - t$ Since Last Observed Transaction: 0.25	1.0	3.0
0.25	$x = 0$	0.851	0.662	0.451
	$x = 1$	0.908	0.648	0.283
	$x = 2$	0.892	0.504	0.100
	$x = 3$	0.873	0.347	0.027
	$x = 4$	0.850	0.211	0.006
	$x = 5$	0.822	0.114	0.001
1.0	$x = 0$	0.662	0.558	0.408
	$x = 1$	0.952	0.809	0.512
	$x = 2$	0.948	0.756	0.325
	$x = 3$	0.944	0.688	0.171
	$x = 4$	0.939	0.607	0.077
	$x = 5$	0.933	0.514	0.031
3.0	$x = 0$	0.451	0.408	0.331
	$x = 1$	0.979	0.916	0.754
	$x = 2$	0.978	0.904	0.676
	$x = 3$	0.977	0.891	0.580
	$x = 4$	0.977	0.875	0.471
	$x = 5$	0.977	0.857	0.359

sponding to $X = 0$, is the probability that someone with no transactions at all in $[0, T]$ is still active at T.

Intuitively, the probability of being active should decrease as the elapsed time $T - t$ since the last transaction increases. Reading across the table horizontally (so keeping X and t constant) it can be seen that this decrease also occurs for the model's computed probability. Similarly, all else being equal the probability of being active should increase as the time t of the last transaction increases. This relationship holds for the Pareto/NBD results, as can be seen by comparing the corresponding entry in cell ($t = 0.25$, $T - t = 1.0$) with that in cell ($t = 1.0$, $T - t = 0.25$).

Looking within each cell at the cases $X = 1, \ldots, 5$ another expected pattern emerges. In columns 2 and 3 ($T - t = 1, 3$) the probability that a customer is active decreases as X increases. For these columns the elapsed time $T - t$ is large compared to the anticipated interpurchase times (based on r/α) of active customers. So if a customer here is still active it is because he or she is waiting an unusually long time before making the next purchase. As the number of observed purchases X made by time t increases, our expectation about this customer's purchase rate (if alive) also increases, and it becomes harder and harder to imagine such an active customer going a period $(t, T]$ without purchasing.

On the other hand, in column 1 the elapsed time $T - t$ is small compared to the anticipated interpurchase time. In this case for a customer still active at T large X-values and the consequent short interpurchase times are quite consistent with the elapsed time $T - t$. As a result, in column 1 the model-based probability that a customer is active does not decrease very greatly with X. So columns 1–3 again follow an expected pattern.

Our final comment regarding the example in Table 1 concerns customers whose number of purchases $X = 0$ in the period $(0, T]$. One might ask "Who is more likely to be still alive: someone who did something ($X > 0$) or someone who did nothing ($X = 0$)?" It seems natural to choose the customer who did something, but the table shows that this need not be the case. For the two cells ($t = 0.25$, $T - t = 1$) and ($t = 0.25$, $T - t = 3$) the individual doing nothing is more likely still active than a customer making one or more purchases. In another cell, ($t = 1$, $T - t = 3$) this $X = 0$ group is more likely active than the heavier purchasers $X = 2, 3, \cdots$ but is less likely active than the light purchasers $X = 1$. As mentioned earlier this result arises because customers with many purchases X in a small period t followed by a long hiatus $T - t$ have probably become inactive. This discussion highlights the inappropriateness of using the hiatus $T - t$ as a heuristic index of customer activity.

Table 1 clarifies our intuitive inferences about customer activity, and suggests that the model works well in cases where intuition can be unreliable. Table 2 clarifies the distinction between choosing "active" customers as above, and choosing those who will be heavy purchasers in a particular future time period. In this case the Pareto/NBD parameters are taken to be $r = 0.415$, $\alpha = 0.415$, $s = 2$ and $\beta = 4$. Results in the table concern customers who made $X = 0, 1, 2, 3, 4$ or 5 transactions in an initial time period $(0, T = 2]$. If $X \geqq 1$ the last transaction in $(0, 2]$ occurred at $t = 1$.

For these customers Table 2 lists the probability of being alive at $T = 2$ and also the expected number of purchases X^* in future period $(T, T + T^*] = (2, 4]$. Both quantities are conditioned on the observed purchase history (X, t, T), so the entries P[Customer Active at $T | r, \alpha, s, \beta, X, t, T$] and $E[X^* | r, \alpha, s, \beta, X, t, T, T^*]$ are calculated using (11)–(13) and (22), respectively. The point of interest is that a customer making zero purchases in $(0, 2]$ is as likely to be still active as one making 4 purchases in $(0, 2]$ with the last at $t = 1$. In both cases the probability is 0.36 that the individual is still active at $T = 2$. However the expected number of transactions in the future period $(2, 4]$ is very different for these two customers. In fact it is over ten times as large for the person making 4 purchases as the one making zero purchases. From (22) it is clear that this is because the expected future purchases depend on both the likelihood of remaining alive, and on the purchase rate *while* the customer is alive. The latter quantity is much greater for the person with 4 transactions.

This illustrates a point made in the last section. A subset of customers can be selected to include those likely to be still active or individuals expected to be heavy buyers next period. Either approach can be appropriate, depending on management objectives, and either can be pursued with the Pareto/NBD. The degree of discrepancy between them,

TABLE 2

The Distinction Between Customers Who Remain Active and Those Who Will be Heavy Purchasers

Purchase Information: X transactions in $(0, T = 2]$ with the last transaction at $t = 1$.

NBD Parameters: $r = 0.415$, $\alpha = 0.415$
Pareto Parameters: $s = 2$, $\beta = 4$

	x					
	0	1	2	3	4	5
P[Customer Active at $T \| r, \alpha, s, \beta, X, t, T$]	0.36	0.60	0.53	0.44	0.36	0.28
$E[X^* \| r, \alpha, s, \beta, X, t, T, T^* = 2]$	0.09	0.53	0.79	0.93	0.99	0.93

in terms of the actual customers chosen, will generally depend on the model parameters (r, α, s, β), the set of purchase histories (X, t, T) observed, and the length T^* of the future period of interest.

These numerical examples were constructed for given values of the model's parameters. We next show how estimates for those four parameters can be obtained.

5. Methods for Parameter Estimation

Three approaches for estimating the Pareto/NBD will be sketched briefly. They involve maximizing the likelihood for observed transaction data, fitting observed moments, or fitting management judgments regarding purchase patterns. The methods differ in the kind of information that is used to describe the purchase/death process. They also differ in statistical efficiency, though this is a more complicated issue.

Maximum Likelihood

Imagine that a random sample with M customers is available, where customer i made $X_i = x_i$ purchases in $(0, T_i]$ with the last transaction at time t_i. The four model parameters (r, α, s, β) can be chosen to maximize the likelihood of seeing the observed data $\{X_i = x_i, t_i, T_i | i = 1, \ldots, M\}$. This likelihood is

$$L(r, \alpha, s, \beta) = \prod_{i=1}^{M} P[X_i = x_i, t_i, T_i | r, \alpha, s, \beta].$$

The probabilities on the right are given by the sum of (A15) and (A16) in the Appendix. For the cases $\alpha > \beta$, $\alpha < \beta$ and $\alpha = \beta$ they can be computed using (A18), (A19), (A23), (A24), (A27), (A28), (A30) and (A31). The principal drawback with this approach is the need to maximize the likelihood using a numerical search algorithm, since an explicit solution for the MLE's ($\hat{r}$, $\hat{\alpha}$, $\hat{s}$, $\hat{\beta}$) has not been obtained.

Fitting Observed Moments

In a given period $(0, T]$ the expected number of purchases made X and the variance of X for randomly chosen customers are given in (17) and (19) as a function of (r, α, s, β). If the purchase information (X_i, t_i, T_i) is available for a random sample of M customers then the corresponding empirical average number of purchases and variance of purchases in $(0, T]$ can also be calculated. Fitting the formulas (17) and (19) to the observed mean and variance for various values of T is another method for estimating the four Pareto/NBD parameters.

To see how the empirical mean and variance are calculated for purchases made in $(0, T]$ note that we can only use customers having $T_i > T$. Let $N(T)$ be the number of customers with $T_i > T$ and $T_{(i)}$ be the ith largest of the $\{T_i | i = 1, \ldots, M\}$. Also let ($X_{(i)}$, $t_{(i)}$) denote the purchase information for the customer with observation time $T_{(i)}$. Then the estimated average number of purchases in $(0, T]$ is

$$\hat{E}[X|T] = \frac{1}{N(T)} \sum_{i=1}^{i_T} X_{(i)}, \tag{24}$$

where i_T is the largest i such that $T_{(i)} \geqq T$. The estimated variance is obtained in the same way:

$$\hat{\mathrm{Var}}\,[X|T] = \frac{1}{N(T) - 1} \sum_{i=1}^{i_T} (X_{(i)} - \hat{E}[X|T])^2. \tag{25}$$

Using (24) and (25) the mean and variance can both be estimated for any positive value of $T \leqq T_{(2)}$. We estimate the parameters (r, α, s, β) by fitting $\hat{E}[X|T]$ and $\hat{\mathrm{Var}}\,[X|T]$ for various values of T using (17) and (19). It is clear that the expectation (17) for varying T-values is a function only of r/α, s and β. So the averages $\hat{E}[X|T]$ can be used to estimate only 3 of the 4 parameters, and $\hat{\text{Var}}\,[X|T]$ must be fit for at least one T-value. Selecting the best $\hat{E}[X|T]$ and $\hat{\text{Var}}\,[X|T]$ values to fit is beyond the scope of this paper. However we will describe a simple 3-step procedure which demonstrates that estimates of (r, α, s, β) can be obtained easily by fitting means and variances.

Step 1. Chose two T-values T_a and T_b such that $T_a \leqq T_{(1)}$, $T_b \leqq T_{(1)}$ and $T_a \neq T_b$. Using (17) we have

$$\frac{E[X|r, \alpha, s, \beta, T_a]}{E[X|r, \alpha, s, \beta, T_b]} = \frac{1 - (\beta/(\beta + T_a))^{s-1}}{1 - (\beta/(\beta + T_b))^{s-1}}. \tag{26}$$

So the proportionate increase in expected purchases as T increases depends only on the two death process parameters s and β. Substituting $\hat{E}[X|T_a]/\hat{E}[X|T_b]$ as an estimate of the left side of (26) gives one equation for estimating (s, β). A second equation is obtained by simply choosing a new pair of T-values T_c and T_d. Thus, estimates for (s, β) are computed by solving numerically the pair of equations:

$$\frac{\hat{E}[X|T_a]}{\hat{E}[X|T_b]} = \frac{1 - (\beta/(\beta + T_a))^{s-1}}{1 - (\beta/(\beta + T_b))^{s-1}}, \tag{27}$$

$$\frac{\hat{E}[X|T_c]}{\hat{E}[X|T_d]} = \frac{1 - (\beta/(\beta + T_c))^{s-1}}{1 - (\beta/(\beta + T_d))^{s-1}}. \tag{28}$$

Step 2. Once $(\hat{s}, \hat{\beta})$ are obtained, estimate r/α by fitting the observed mean using (17) for some time period $T_e \leqq T_{(1)}$. This results in the estimate for r/α:

$$\left(\frac{\hat{r}}{\alpha}\right) = \frac{(\hat{s} - 1)}{\hat{\beta}} \frac{\hat{E}[X|T_e]}{1 - (\hat{\beta}/(\hat{\beta} + T_e))^{\hat{s}-1}}. \tag{29}$$

Step 3. Substitute $\hat{s}$, $\hat{\beta}$ and the estimated ratio $(\widehat{r/\alpha})$ in (19) to fit the observed variance $\hat{\text{Var}}\,[X|T_f]$ for some $T_f \leqq T_{(2)}$. With these substitutions, (19) can be solved for the fourth parameter $\hat{\alpha}$.

Although the selection of good values for T_a through T_f is deferred to future work, we note that a reasonable choice for T_e and T_f would be to maximize the "total time on test," i.e. the total time available across customers for estimating the mean and variance. This would involve choosing $T_e = T_f \equiv T^* \leqq T_{(2)}$ to maximize

$$\text{Total Time on Test} = \sum_{i=1}^{i_{T^*}} T^*, \tag{30}$$

where as above i_{T^*} is the largest i such that $T_{(i)} \geqq T^*$.

Estimation Using Management Judgments

In some cases it would be desirable to use the Pareto/NBD model before any customer purchase data are available. For example, imagine a firm that wishes to set a policy for retaining customers on a newly constructed or acquired mailing list. Eventually, enough purchase information will flow in to estimate (r, α, s, β) but we want a policy to use until then.

In this situation a series of structured subjective judgements made by management can be used to obtain $(\hat{r}, \hat{\alpha}, \hat{s}, \hat{\beta})$. The firm above would probably find it easiest to give subjective estimates of $\hat{E}[X|T]$ for varying T-values. As in (27)–(29) this can generate estimates of $\hat{s}$, $\hat{\beta}$ and $(\widehat{r/\alpha})$. The fourth parameter can be obtained by fitting the judged probability that fewer than $X = x$ purchases occur in $(0, T]$ (using (A40), (A43) and (A45)). Alternatively, one can fit the judged probability that fewer than $X = x$ purchases occur in $(0, T]$ for a customer *still alive at T* (using (6)). These judgments about probabilities will typically be easier to make than an assessment of the variance $\hat{\mathrm{Var}}\,[X|T]$.

6. Discussion

In this paper we have discussed the identification of active customers using a relatively simple yet realistic model. Two general conditions are required for the approach to be appropriate. First, the time at which customers become inactive must be unknown to the analyst, or difficult to know. If that time *is* observed, the questions posed in this paper can be answered by direct examination without our model. In that case it would be more interesting to identify factors determining the (observed) time of death, using the models described in Heckman and Singer (1982) and Trussell and Richards (1985).

Second, the Pareto/NBD assumptions imply that any number of purchases can be made at any time, and the customer could become inactive at any time. That is, there are no discrete, observed, periodic purchase opportunities which are required in order to make a transaction. So, for example, our model should not represent the viewership of a new television series, since here a "purchase" (viewing the program) can only occur during those periodic opportunities when the program is aired.

Table 3 lists some processes that we believe do usually meet these two general conditions (top left cell) and some that do not. The Pareto/NBD is inappropriate for examining renewable service contracts, HMO membership and subscriptions to a particular magazine, since the opportunities for transactions occur at regular, observed intervals. While the model is more appropriate for processes like usage of a bank account or telephone service (since the purchases can occur continuously) it isn't usually needed since the firm knows exactly when the service is terminated. However, one exception is the identification of bank accounts which, unbeknownst to the bank, have become permanently inactive. (This can occur through a customer's forgetting or negligence. It also can happen when a customer dies and the heirs and executor are

TABLE 3
Scenarios for Analyzing Active vs. Inactive Customers

		Time At Which Customers Become Inactive	
		Unobserved	Observed
Opportunities For Transactions	Continuous/ Unobserved	—brokerage transactions —purchases of a consumer packaged good —department store credit card usage —doctor or dentist visits —mail catalog sales —pregnancies/birth control	—telephone service (local and long distance) —cable television service —bank accounts
	Discrete/ Observed	—church attendance —vendor sales of magazines —television viewing	—renewable service contracts —magazine subscriptions —HMO membership

unaware of the account.) It is a common enough occurrence to be a concern for some banks and state legislatures.

For processes like church attendance and television viewing the opportunities for a transaction occur regularly, so our model is again inappropriate. But since the organization does not typically know when a customer here becomes inactive, counting and identifying active customers is still a nontrivial problem. We have developed an alternate model that can be used for processes in this bottom left cell of Table 3, which will be the subject of some future work.

Finally, a variety of processes (indicated in the table's top left cell) fit the general criteria for the Pareto/NBD, including brokerage transactions and purchases of a consumer packaged good. Our model could also be used to assess the effectiveness of a government birth control program. By discussing this application here we will indicate the wide applicability of our general approach—while at the same time highlighting some potential problems with the model's specific assumptions.

Obviously, we could model a fertile woman as a Poisson process with respect to the event of conception if we delete the 9-month gestation periods from the time axis. That is, each month that the woman is not pregnant there is a constant probability that she will become pregnant. If the woman is not using any birth control procedure, we could term her as being "alive." If she started using some effective birth control method, we could say that she has become "dead." Hence, if we saw a woman who has 4 children between ages 18 and 22 and had no additional children by age 30, we could assume a fairly high probability that some form of birth control (dead) had occurred. Obviously, across women there will be variations in pregnancy rates with no birth control (the λ's) and in the receptivity to a birth control procedure (the μ's). All of the above implies that our model may be a starting point for assessing the use of birth control—but clearly there are major problems.

Birth control procedures are never 100 percent effective—hence, there is no pure death. Our model allows no resurrection from death, yet a woman using birth control can stop using it and become "alive" again. The exponential time until the use of birth control may not be realistic if the country or region varies the pressures or incentives over time. Finally, the women's individual level conception rates (the λ's) will vary as they age and become zero at some future point in time.

A little reflection will indicate that all of the above inadequacies of our modelling of the birth control problem apply in varying degrees to alive and dead customers. Customers can come back from the dead—witness the catalog companies sending out their multiple "absolutely last chance" reminders. Lifestyle changes can alter individuals' purchasing rates. And if lifestyle changes can cause "death," we know that at least some of these changes do not occur randomly. In short, our model will often capture the spirit, but not the complete essence, of the counting the number of active customers problem. Therefore we should conclude this discussion by restating what we have accomplished and listing further research that needs to be done.

What Have We Done?

Hopefully, we have accomplished two things in this article. First, we have tried to convince the reader that determining which customers are active is an important, widespread problem. Second, we developed the most mathematically tractable—and still realistic—model of the problem and solved the conditional probability $P(\text{Alive}|\text{Information})$. Now for what remains to be done.

Parameter Estimation

Our model has 4 parameters—two each for the gamma mixing distributions on λ and μ. As §5 indicated, more work is needed to determine the effectiveness of various methods for estimating these parameters.

Probabilistic Extensions

The robustness of the model to deviations from our assumptions should be assessed. This would certainly include examining the effects of correlations between λ and μ, nongamma mixing distributions, nonexponential interevent times and perhaps the relaxation of "death" as an absorbing state. An alternative to exponential interevent times has been developed by Chatfield and Goodhardt (1973), Morrison and Schmittlein (1981) and Schmittlein and Morrison (1983).

It would also be very appealing to extend this general approach by modelling the distribution for λ and μ as a function of demographic or product usage variables. This is easy to do in principle. A natural choice would be a loglinear relationship between the expected value of λ (or μ) and the covariates $d_1, \ldots, d_D$, i.e.,

$$\log(E[\lambda]) = \log(r/\alpha) = \gamma_0 + \gamma_1 d_1 + \cdots + \gamma_D d_D \tag{31}$$

and similarly for μ. The logarithmic transformations are useful to reflect the restrictions $\lambda \geqq 0$ and $\mu \geqq 0$. Some of the covariates d_i may also be time-varying, to capture the effect of some exogenous event. For example if our brokerage firm were acquired by a financial services conglomerate, purchase and/or death rates could change simultaneously for many customers.[1] With extensions like (31) relaxation of the gamma mixing distribution assumptions may be particularly important in light of recent work by Heckman and Singer (1982) and Trussell and Richards (1985). They vary the functional form assumed for an unobservable variable and find a noticeable effect on the parameter estimates in regression models for duration times.

Methodological Issues

Should the analyst use all of the available data? Consider a customer who has been doing business on a regular basis with Merrill Lynch for 40 years. If this complete past history were put into our model the Bayesian updating on this customer's death rate μ would drive our expected value of his or her particular μ close to zero. That is, we would expect the customer to be "alive" for a very long time into the future. If nothing else, this is ignoring the actuarial tables. The implicit stationarity of our model will be strained by very long histories. Perhaps we should use (say) only two years of data even if we have more.

Should the first purchase count? Some situations have forced trial. Even if the first purchase is done "normally," there is an old adage that "advertising creates trial, but the product delivers repeat." Clearly, some judgment is required in determining how to count the first purchase.

Qualitative Insights from Quantitative Models

In some earlier work done at Merrill Lynch by Morrison, Chen, Karpis and Britney (1982) the stability of the better clients was investigated. The firm wished to know how likely it would be for a top customer this year to be a big revenue producer next year and on into the future. Obviously management had some data and good intuition as to year to year changes. However, it was important to know how many current big customers were still going to be big customers 5 to 7 years down the road. Half of them? Ten percent? A quantitative model was required to answer this qualitative question.

In a similar vein, consider the hypothetical executive in our Introduction. This individual does not need to know whether the percentage of active customers on the list of all past customers is 83 vs. 85 percent. But he or she surely would like to know whether the correct percentage is 50, 75 or 90. Even if the assumptions of our model are

[1] This example was pointed out by an anonymous reviewer, to whose brokerage firm it in fact occurred.

not precisely correct, this model presents an easily implemented approach to determining a qualitative assessment of the size of the currently active customer pool, and the rate at which that pool's size is increasing or decreasing. As our numerical examples showed, one's intuition on the probability that a particular event history represents an active customer is not always very good. Our simple model can help guide the manager's intuition.[2]

[2] The authors would like to thank Erin Anderson for helpful comments. This work was supported by a grant from the Center for Marketing Strategy Research, The Wharton School, University of Pennsylvania.

Appendix

1. *Derivation of P[Alive|Information] at the Individual Level*

Here we derive the probability that the time of death τ occurs after the observation period $(0, T]$ for a person with known (Poisson) purchase rate λ and (exponential) death rate μ. This probability is conditioned on the information that $X = x > 0$ transactions occurred in $(0, T]$ at times $0 < \zeta_1 \leqq \zeta_2 \leqq \cdots \leqq \zeta_x \equiv t \leqq T$. Denote the jth interpurchase time by $\xi_j \equiv \zeta_j - \zeta_{j-1}; j = 1, \ldots, x$; where $\zeta_0 \equiv 0$. Also, denote the event that no purchase occurred in the interval $(t, T]$ by ϕ_{T-t}. Then the vector $\xi \equiv (\xi_1, \cdots \xi_x, \phi_{T-t})$ represents all of the available data for this customer: the x interpurchase times, and the observed interval after the last purchase where no transactions occurred. With these definitions $\sum_{j=1}^{x} \xi_j = t$.

The probability that a customer with transaction pattern ξ is still alive at T can be written using Bayes theorem as

$$P[\tau > T|\lambda, \mu, \xi] = \frac{f_1[\xi|\lambda, \mu, \tau > T]P[\tau > T|\lambda, \mu]}{g[\xi|\lambda, \mu]} \tag{A1}$$

$$= \frac{f_1[\xi|\lambda, \mu, \tau > T]P[\tau > T|\lambda, \mu]}{f_1[\xi|\lambda, \mu, \tau > T]P[\tau > T|\lambda, \mu] + f_2[\xi, t < \tau \leqq T|\lambda, \mu]} \tag{A2}$$

where f_1, f_2 and g are the likelihoods of seeing the events indicated. Using equation (2),

$$P[\tau > T|\lambda, \mu] = e^{-\mu T}. \tag{A3}$$

f_1 and f_2 in (A2) can be rewritten as

$$f_1[\xi|\lambda, \mu, \tau > T] = f_3[\xi|\zeta_x = t, \phi_{T-t}, \lambda, \mu, \tau > T]\, f_4\, [\zeta_x = t, \phi_{T-t}|\lambda, \mu, \tau > T] \quad \text{and} \tag{A4}$$

$$f_2[\xi, t < \tau \leqq T|\lambda, \mu] = f_5[\xi|\zeta_x = t, \phi_{\tau-t}, \lambda, \mu]\, f_6\, [\zeta_x = t, \phi_{\tau-t}, t < \tau \leqq T|\lambda, \mu] \tag{A5}$$

where again f_3 through f_6 represent the likelihoods of the indicated events.

Conditioned on the time of the last Poisson event $\zeta_x = t$, the earlier $x - 1$ transactions randomly split the available time line, i.e., they are independent uniform random variables on $(0, t]$. So, conditioned on $\zeta_x = t$, each of the $x - 1$ values $\zeta_j, j = 1, \ldots, x - 1$, has likelihood $1/t$ (i.e., the uniform p.d.f. on $(0, t]$). Consequently

$$f_3[\xi|\zeta_x = t, \phi_{T-t}, \lambda, \mu, \tau > T] = f_5[\xi|\zeta_x = t, \phi_{\tau-t}, t < \tau \leqq T, \lambda, \mu] = t^{-(x-1)}. \tag{A6}$$

Conditioned on $\tau > T$, the events $\zeta_x = t$ and ϕ_{T-t} are independent. Since ζ_x is the sum of x i.i.d. exponential interpurchase times, each with mean $1/\lambda$, it has a gamma distribution with p.d.f. (Meyer 1970, p. 221):

$$f(\zeta_x|x, \lambda) = \frac{\lambda}{\Gamma(x)}(\lambda\zeta_x)^{x-1}e^{-\lambda\zeta_x}. \tag{A7}$$

Using (A7), f_4 becomes

$$f_4\, [\zeta_x = t, \phi_{T-t}|\lambda, \mu, \tau > T] = \frac{\lambda^x t^{x-1}}{\Gamma(x)} e^{-\lambda t}[e^{-\lambda(T-t)}] = \frac{\lambda^x t^{x-1} e^{-\lambda T}}{\Gamma(x)}. \tag{A8}$$

Similarly, f_6 is

$$f_6\, [\zeta_x = t, \phi_{\tau-t}, t < \tau \leqq T|\lambda, \mu] = \int_t^T f_4\, [\zeta_x = t, \phi_{y-t}|\lambda, \mu, y > t] f\, [\tau = y|\lambda, \mu] dy$$

$$= \text{, using (A7) and (2),}$$

$$= \int_t^T \frac{\lambda^x t^{x-1} e^{-\lambda y}}{\Gamma(x)} \mu e^{-\mu y} dy$$

$$= \frac{\lambda^x t^{x-1}}{\Gamma(x)} \left(\frac{\mu}{\lambda+\mu}\right)[e^{-(\lambda+\mu)t} - e^{-(\lambda+\mu)T}]. \tag{A9}$$

Substituting (A6), (A8) and (A9) in (A4) and (A5), and these into (A2) yields the desired result

$$P[\tau > T|\lambda, \mu, \xi] = \frac{1}{1 + \left(\frac{\mu}{\lambda+\mu}\right)[e^{(\lambda+\mu)(T-t)} - 1]}. \tag{A10}$$

Note that the right side of (A10) depends only on $(X = x, t, T)$ and not on all of the original information in ξ. For this reason in the text we write the left side of (A10) as $P[\tau > T]\lambda, \mu, X = x, t, T]$, i.e., the individual-level probability that a customer is still alive at T given (λ, μ) and the information $(X = x, t, T)$.

2. *Derivation of P[Alive|Information] for a Randomly Chosen Customer*

We turn now to the probability that the time of death τ occurs after the observation period $(0, T]$ for a randomly chosen individual whose purchase rate λ and death rate μ are unknown. Conditioned on the customer's making $X = x > 0$ transactions during $(0, T]$ with the last at $t < T$, the probability that this person is still alive at T is the expectation of the individual-level result (A10) over the updated distribution for λ and μ:

$$P[\tau > T|r, \alpha, s, \beta, X = x, t, T]$$

$$= \int_0^\infty \int_0^\infty P[\tau > T|\lambda, \mu, X = x, t, T] f(\lambda, \mu|r, \alpha, s, \beta, X = x, t, T) d\lambda d\mu. \tag{A11}$$

The conditional distribution of (λ, μ) in (A11) can be written using Bayes theorem as

$$f(\lambda, \mu|r, \alpha, s, \beta, X = x, t, T) = \frac{P[X = x, t, T|\lambda, \mu] g(\lambda|r, \alpha) h(\mu|s, \beta)}{\int_0^\infty \int_0^\infty P[X = x, t, T|\lambda, \mu] g(\lambda|r, \alpha) h(\mu|s, \beta) d\lambda d\mu} \tag{A12}$$

where $g(\ )$ and $h(\ )$ are the prior distribution p.d.f.'s given in (4) and (5) respectively. Substituting (A12) and (A1) in (A11) we have

$$P[\tau > T|r, \alpha, s, \beta, X = x, t, T] = \frac{\int_0^\infty \int_0^\infty P[X = x, t, T|\lambda, \mu, \tau > T] P[\tau > T|\lambda, \mu] g(\lambda|r, \alpha) h(\mu|s, \beta) d\lambda d\mu}{\int_0^\infty \int_0^\infty P[X = x, t, T|\lambda, \mu] g(\lambda|r, \alpha) h(\mu|s, \beta) d\lambda d\mu}. \tag{A13}$$

By conditioning the denominator of (A13) on the two possible cases $\tau > T$ and $t < \tau < T$, (A13) can be rewritten as

$$P[\tau > T|r, \alpha, s, \beta, X = x, t, T] = \frac{A}{A+B} = \frac{1}{1 + (B/A)} \quad \text{where} \tag{A14}$$

$$A = \int_0^\infty \int_0^\infty P[X = x, t, T|\lambda, \mu, \tau > T] P[\tau > T|\lambda, \mu] g(\lambda|r, \alpha) h(\mu|s, \beta) d\lambda d\mu \quad \text{and} \tag{A15}$$

$$B = \int_0^\infty \int_0^\infty P[X = x, t, T|\lambda, \mu, t < \tau < T] P[t < \tau < T|\lambda, \mu] g(\lambda|r, \alpha) h(\mu|s, \beta) d\lambda d\mu. \tag{A16}$$

Substituting (4), (5), (A3) and (A8) in (A15) yields

$$A = \frac{t^{x-1}}{\Gamma(x)} \int_0^\infty \int_0^\infty \lambda^x e^{-\lambda T} e^{-\mu T} \frac{\alpha^r \lambda^{r-1}}{\Gamma(r)} e^{-\alpha\lambda} \frac{\beta^s \mu^{s-1}}{\Gamma(s)} e^{-\beta\mu} d\lambda d\mu$$

$$= \frac{t^{x-1} \alpha^r \beta^s \Gamma(r+x)}{\Gamma(x)\Gamma(r)(\alpha+T)^{r+x}(\beta+T)^s} \left[\int_0^\infty \frac{(\alpha+T)^{r+x}\lambda^{r+x-1}}{\Gamma(r+x)} e^{-(\alpha+T)\lambda} d\lambda\right] \left[\int_0^\infty \frac{(\beta+T)^s \mu^{s-1}}{\Gamma(s)} e^{-(\beta+T)\mu} d\mu\right]. \tag{A17}$$

Both integrals in (A17) equal 1, being the integral of a gamma p.d.f. over the domain $(0, \infty)$. So A becomes

$$A = t^{x-1} \frac{\alpha^r}{(\alpha+T)^{r+x}} \frac{\beta^s}{(\beta+T)^s} \frac{\Gamma(r+x)}{\Gamma(r)\Gamma(x)}. \tag{A18}$$

Substituting (2), (4), (5) and (A9) in (A16) yields,

$$B = V_1 - V_2 \quad \text{where} \tag{A19}$$

$$V_1 = \frac{t^{x-1}}{\Gamma(x)} \int_0^\infty \int_0^\infty \lambda^x \left(\frac{\mu}{\lambda+\mu}\right) e^{-(\lambda+\mu)t} \frac{\alpha^r \lambda^{r-1}}{\Gamma(r)} e^{-\alpha\lambda} \frac{\beta^s \mu^{s-1}}{\Gamma(s)} e^{-\beta\mu} d\lambda d\mu \quad \text{and} \tag{A20}$$

$$V_2 = \frac{t^{x-1}}{\Gamma(x)} \int_0^\infty \int_0^\infty \lambda^x \left(\frac{\mu}{\lambda+\mu}\right) e^{-(\lambda+\mu)T} \frac{\alpha^r \lambda^{r-1}}{\Gamma(r)} e^{-\alpha\lambda} \frac{\beta^s \mu^{s-1}}{\Gamma(s)} e^{-\beta\mu} d\lambda d\mu. \tag{A21}$$

To evaluate V_1 and V_2 it is convenient to consider separately the three cases $\alpha > \beta$, $\alpha < \beta$ and $\alpha = \beta$.

Case 1: $\alpha > \beta$. Making the change of variables $z = \lambda + \mu$ and $p = \mu/(\lambda + \mu)$ in (A20) V_1 is

$$V_1 = \frac{t^{x-1}\alpha^r\beta^s}{\Gamma(x)\Gamma(r)\Gamma(s)} \int_0^1 \int_0^\infty p^s(1-p)^{r+x-1} z^{r+x+s-1} e^{-(\alpha+t-(\alpha-\beta)p)z} dz dp$$

$$= \frac{t^{x-1}\alpha^r\beta^s\Gamma(r+x+s)}{\Gamma(x)\Gamma(r)\Gamma(s)} \int_0^1 p^s(1-p)^{r+x-1}[\alpha + t - (\alpha-\beta)p]^{-(r+x+s)} dp$$

$$= \frac{t^{x-1}\alpha^r\beta^s\Gamma(r+x+s)}{\Gamma(x)\Gamma(r)\Gamma(s)(\alpha+t)^{r+x+s}} \int_0^1 p^s(1-p)^{r+x-1}\left[1 - \left(\frac{\alpha-\beta}{\alpha+t}\right)p\right]^{-(r+x+s)} dp. \tag{A22}$$

The integral in (A22) can be written as a Gauss hypergeometric function (Abramowitz and Stegun 1972, p. 558) so V_1 becomes

$$V_1 = \frac{t^{x-1}\alpha^r\beta^s\Gamma(r+x+s)}{\Gamma(x)\Gamma(r)\Gamma(s)(\alpha+t)^{r+x+s}} \frac{\Gamma(s+1)\Gamma(r+x)}{\Gamma(r+x+s+1)} F\left(r+x+s, s+1; r+x+s+1; \frac{\alpha-\beta}{\alpha+t}\right)$$

$$= \frac{t^{x-1}\alpha^r\beta^s\Gamma(r+x)}{\Gamma(x)\Gamma(r)(\alpha+t)^{r+x+s}} \left(\frac{s}{r+x+s}\right) F\left(r+x+s, s+1; r+x+s+1; \frac{\alpha-\beta}{\alpha+t}\right). \tag{A23}$$

The integral for V_2 in (A21) is evaluated by making the same change of variables as for V_1, with the result

$$V_2 = \frac{t^{x-1}\alpha^r\beta^s\Gamma(r+x)}{\Gamma(x)\Gamma(r)(\alpha+T)^{r+x+s}} \left(\frac{s}{r+x+s}\right) F\left(r+x+s, s+1; r+x+s+1; \frac{\alpha-\beta}{\alpha+T}\right). \tag{A24}$$

For this case $\alpha > \beta$, substituting (A24), (A23), (A19) and (A18) in (A14) gives the desired result:

$$P[\tau > T | r, \alpha, s, \beta, X = x, t, T] = \left\{1 + \frac{s}{r+x+s}\left[\left(\frac{\alpha+T}{\alpha+t}\right)^{r+x}\left(\frac{\beta+T}{\alpha+T}\right)^s \times F(a_1, b_1; c_1; z_1(t)) - \left(\frac{\beta+T}{\alpha+T}\right)^s F(a_1, b_1; c_1; z_1(T))\right]\right\}^{-1} \quad \text{where} \tag{A25}$$

$$a_1 = r+x+s; \quad b_1 = s+1; \quad c_1 = r+x+s+1; \quad z_1(y) = \frac{\alpha-\beta}{\alpha+y}.$$

Case 2. $\alpha < \beta$. Making the change of variables $z = \lambda + \mu$ and $q = \lambda/(\lambda + \mu)$ in (A20), V_1 is

$$V_1 = \frac{t^{x-1}\alpha^r\beta^s}{\Gamma(x)\Gamma(r)\Gamma(s)} \int_0^1 \int_0^\infty q^{r+x-1}(1-q)^s z^{r+x+s-1} e^{-(\beta+t-(\beta-\alpha)q)z} dz dq \tag{A26}$$

$$= \frac{t^{x-1}\alpha^r\beta^s\Gamma(r+x+s)}{\Gamma(x)\Gamma(r)\Gamma(s)(\beta+t)^{r+x+s}} \int_0^1 q^{r+x-1}(1-q)^s\left[1 - \left(\frac{\beta-\alpha}{\beta+t}\right)q\right]^{-(r+x+s)} dq$$

$$= \frac{t^{x-1}\alpha^r\beta^s\Gamma(r+x)}{\Gamma(x)\Gamma(r)(\beta+t)^{r+x+s}} \left(\frac{s}{r+x+s}\right) F\left(r+x+s, r+x; r+x+s+1; \frac{\beta-\alpha}{\beta+t}\right). \tag{A27}$$

Using the same change of variables in (A21), V_2 becomes

$$V_2 = \frac{t^{x-1}\alpha^r\beta^s\Gamma(r+x)}{\Gamma(x)\Gamma(r)(\beta+T)^{r+x+s}} \left(\frac{s}{r+x+s}\right) F\left(r+x+s, r+x; r+x+s+1; \frac{\beta-\alpha}{\beta+T}\right). \tag{A28}$$

Substituting (A28), (A27), (A19) and (A18) in (A14) gives the desired result for this case where $\alpha < \beta$:

$$P[\tau > T | r, \alpha, s, \beta, X = x, t, T] = \left\{1 + \frac{s}{r+x+s}\left[\left(\frac{\alpha+T}{\beta+t}\right)^{r+x}\left(\frac{\beta+T}{\beta+t}\right)^s F(a_2, b_2; c_2; z_2(t)) - \left(\frac{\alpha+T}{\beta+T}\right)^{r+x} F(a_2, b_2; c_2; z_2(T))\right]\right\}^{-1} \quad \text{where} \tag{A29}$$

$$a_2 = r+x+s; \quad b_2 = r+x; \quad c_2 = r+x+s+1; \quad z_2(y) = \frac{\beta-\alpha}{\beta+y}.$$

Case 3. $\alpha = \beta$. Substituting $\alpha = \beta$ in (A26), V_1 becomes

$$V_1 = \frac{t^{x-1}\alpha^{r+s}\Gamma(r+x+s)}{\Gamma(x)\Gamma(r)\Gamma(s)(\alpha+t)^{r+x+s}} \int_0^1 p^s(1-p)^{r+x-1}dp$$

$$= \frac{t^{x-1}\alpha^{r+s}\Gamma(r+x+s)}{\Gamma(x)\Gamma(r)\Gamma(s)(\alpha+t)^{r+x+s}} \frac{\Gamma(s+1)\Gamma(r+x)}{\Gamma(r+x+s+1)} = \frac{t^{x-1}\alpha^{r+s}\Gamma(r+x)}{\Gamma(x)(\alpha+t)^{r+x+s}\Gamma(r)}\left(\frac{s}{r+x+s}\right). \tag{A30}$$

Similarly, V_2 becomes

$$V_2 = \frac{t^{x-1}\alpha^{r+s}\Gamma(r+x)}{\Gamma(x)(\alpha+T)^{r+x+s}\Gamma(r)}\left(\frac{s}{r+x+s}\right). \tag{A31}$$

Substituting (A30), (A31), (A19) and (A18) in (A14) yields the desired probability that the customer is still active when $\alpha = \beta$:

$$P[\tau > T \mid r, \alpha, s, \beta, X = x, t, T] = \frac{1}{1 + (s/(r+x+s))[((\alpha+T)/(\alpha+t))^{r+x+s} - 1]}. \tag{A32}$$

As noted above, these derivations of P[Alive|Information] at both the individual level (A10) and the aggregate level (A25), (A29) and (A32) assume that $X \geqq 1$; i.e. that at least one transaction was made in $(0, T]$. (Hence it makes sense to define a quantity t = time of the last purchase in $(0, T]$.) The desired probabilities for the special case $X = 0$ can be derived separately. We will not do so here since the derivation follows the same steps as in (A1)–(A32). The result is that the probabilities when $X = 0$ are obtained by substituting ($X = 0$, $t = 0$) in equations (A10), (A25), (A29) and (A32).

3. *Probability Distribution for the Number of Purchases Made in* $(0, T]$

We derive here the distribution for the number of purchases made by a randomly chosen customer following the Pareto/NBD model. This probability of observing $X = x$ purchases in time period $(0, T]$ can be obtained by conditioning on whether the individual is still active at T. If so ($\tau > T$) then the distribution for the number of purchases in $(0, T]$ is the NBD as in (6). If not, purchases follow the NBD (6) for the period $(0, \tau]$ where $\tau < T$. Using this conditioning idea the desired probability of observing $X = x$ events in $(0, T]$ is

$$P[X = x \mid r, \alpha, s, \beta, T] = P[X = x \mid r, \alpha, \tau > T]P[\tau > T \mid s, \beta] + \int_0^T P[X = x \mid r, \alpha, \tau > \theta] f(\theta \mid s, \beta) d\theta \tag{A33}$$

where $f(\ )$ is the p.d.f. for the duration time τ as in (9).

Substituting (6) and (9) in (A33) yields

$$P[X = x \mid r, \alpha, s, \beta, T] = \binom{x+r-1}{x}\left(\frac{\alpha}{\alpha+T}\right)^r\left(\frac{T}{\alpha+T}\right)^x\left(\frac{\beta}{\beta+T}\right)^s + \binom{x+r-1}{x} s\alpha^r\beta^s V_3 \quad \text{where} \tag{A34}$$

$$V_3 = \int_0^T \theta^x(\alpha+\theta)^{-(r+x)}(\beta+\theta)^{-(s+1)}d\theta. \tag{A35}$$

To evaluate V_3 it is helpful again to consider the cases $\alpha > \beta$, $\alpha < \beta$ and $\alpha = \beta$ separately.

Case 1. $\alpha > \beta$. Making the change of variable $y = \alpha + \theta$, V_3 becomes

$$V_3 = \int_\alpha^{\alpha+T} (y-\alpha)^x y^{-(r+x)}(y-\alpha+\beta)^{-(s+1)}dy \tag{A36}$$

$$= \sum_{j=0}^{x}\binom{x}{j}(-\alpha)^j \int_\alpha^{\alpha+T} y^{-(r+j)}(y-\alpha+\beta)^{-(s+1)}dy$$

$$= \sum_{j=0}^{x}\binom{x}{j}(-\alpha)^j \left[\int_\alpha^{\infty} y^{-(r+j)}(y-\alpha+\beta)^{-(s+1)}dy - \int_{\alpha+T}^{\infty} Y^{-(r+j)}(y-\alpha+\beta)^{-(s+1)}dy\right]. \tag{A37}$$

Making the change of variable $z = \alpha/y$ in the first integral and $z = (\alpha + T)/y$ in the second integral of (A37), V_3 becomes

$$V_3 = \sum_{j=0}^{x}\binom{x}{j}(-\alpha)^j \Bigg[\alpha^{-(r+j+s)}\int_0^1 z^{r+j+s-1}$$

$$\times\left[1-\left(\frac{\alpha-\beta}{\alpha}\right)z\right]^{-(s+1)}dz - (\alpha+T)^{-(r+j+s)}\int_0^1 z^{r+j+s-1}\left[1-\left(\frac{\alpha-\beta}{\alpha+T}\right)z\right]^{-(s+1)}dz\Bigg]. \tag{A38}$$

The integrals in (A38) can be expressed as Gauss hypergeometric functions (Abramowitz and Stegun 1972, p. 558), yielding

$$V_3 = \sum_{j=0}^{x} \binom{x}{j} (-\alpha)^j \left[\frac{\alpha^{-(r+j+s)}}{r+j+s} F\left(s+1, r+j+s; r+j+s+1; \frac{\alpha-\beta}{\alpha}\right) - \frac{(\alpha+T)^{-(r+j+s)}}{r+j+s} F\left(s+1, r+j+s; r+j+s+1; \frac{\alpha-\beta}{\alpha+T}\right)\right]. \quad \text{(A39)}$$

Substituting (A39) in (A34) gives the desired probability distribution for the case $\alpha > \beta$:

$$P[X = x|r, \alpha, s, \beta, T] = \binom{x+r-1}{x} \left(\frac{\alpha}{\alpha+T}\right)^r \left(\frac{T}{\alpha+T}\right)^x \left(\frac{\beta}{\beta+T}\right)^s + \binom{x+r-1}{x} \sum_{j=0}^{x} \binom{x}{j} (-\alpha)^j \left(\frac{s}{s+r+j}\right) \alpha^r \beta^s \left[\alpha^{-(r+j+s)} F\left(s+1, r+j+s; r+j+s+1; \frac{\alpha-\beta}{\alpha}\right) - (\alpha+T)^{-(r+j+s)} F\left(s+1, r+j+s; r+j+s+1; \frac{\alpha-\beta}{\alpha+T}\right)\right]. \quad \text{(A40)}$$

Case 2: $\alpha < \beta$. Making the change of variable $y = \beta + \theta$ in (A35), V_3 becomes

$$V_3 = \int_{\beta}^{\beta+T} (y-\beta)^x y^{-(s+1)} (y-\beta+\alpha)^{-(r+x)} dy. \quad \text{(A41)}$$

The integral in (A41) has the same form as that in (A36), with α and β changing roles, and also $(s+1)$ and $(r+x)$ changing places. So using the solution (A39) V_3 is

$$V_3 = \sum_{j=0}^{x} \binom{x}{j} (-\beta)^j \left[\frac{\beta^{-(r+j+s)}}{r+j+s} F\left(r+x, r+j+s; r+j+s+1; \frac{\beta-\alpha}{\beta}\right) - \frac{(\beta+T)^{-(r+j+s)}}{r+j+s} F\left(r+x, r+j+s; r+j+s+1; \frac{\beta-\alpha}{\beta+T}\right)\right]. \quad \text{(A42)}$$

Substituting (A42) in (A34) gives the desired probability distribution for the case $\alpha < \beta$:

$$P[X = x|r, \alpha, s, \beta, T] = \binom{x+r-1}{x} \left(\frac{\alpha}{\alpha+T}\right)^r \left(\frac{T}{\alpha+T}\right)^x \left(\frac{\beta}{\beta+T}\right)^s + \binom{x+r-1}{x} \sum_{j=0}^{x} \binom{x}{j} (-\beta)^j \left(\frac{s}{s+r+j}\right) \alpha^r \beta^s \left[\beta^{-(r+j+s)} F\left(r+x, r+j+s; r+j+s+1; \frac{\beta-\alpha}{\beta}\right) - (\beta+T)^{-(r+j+s)} F\left(r+x, r+j+s; r+j+s+1; \frac{\beta-\alpha}{\beta+T}\right)\right]. \quad \text{(A43)}$$

Case 3: $\alpha = \beta$. Substituting $\alpha = \beta$ in (A38) yields

$$V_3 = \sum_{j=0}^{x} \binom{x}{j} (-\alpha)^j \left[\frac{\alpha^{-(r+j+s)}}{r+j+s} - \frac{(\alpha+T)^{-(r+j+s)}}{r+j+s}\right]. \quad \text{(A44)}$$

Substituting (A44) in (A34) gives the probability distribution for purchases in $(0, T]$ for the case $\alpha = \beta$:

$$P[X = x|r, \alpha, s, \beta, T] = \binom{x+r-1}{x} \left(\frac{\alpha}{\alpha+T}\right)^{r+s} \left(\frac{T}{\alpha+T}\right)^x + \binom{x+r+1}{x} \sum_{j=0}^{x} \binom{x}{j} (-\alpha)^j \alpha^{r+s} \left(\frac{s}{s+r+j}\right) [\alpha^{-(r+j+s)} - (\alpha+T)^{-(r+j+s)}]. \quad \text{(A45)}$$

References

Abramowitz, M. and I. A. Stegun (eds.) *Handbook of Mathematical Functions*, Dover Publications Inc., New York, 1972.

Chatfield, C. and G. J. Goodhardt, "A Consumer Purchasing Model with Erlang Inter-Purchase Times," *J. Amer. Statist. Assoc.*, 68 (1973), 828–835.

EHRENBERG, A. S. C., *Repeat Buying*, North-Holland, Amsterdam, 1972.

ESKIN, G. J., "Dynamic Forecasts of New Product Demand Using a Depth of Repeat Model," *J. Marketing Res.*, 10 (1973), 115–129.

——— AND J. MALEC, "A Model for Estimating Sales Potential Prior to the Test Market," *Proc., Educators Conference, Ser. No.* 39, American Marketing Association, Chicago, 1976, 220–233.

FELLER, W., *An Introduction to Probability Theory and Its Applications. Volume II* (2nd ed.), John Wiley, New York, 1971.

HECKMAN, J. AND B. SINGER, "Population Heterogeneity in Demographic Models," in K. Land and A. Rogers (eds.), *Multidimensional Mathematical Demography*, Academic Press, New York, 1982.

JOHNSON, N. L. AND S. KOTZ, *Continuous Univariate Distributions*—1, John Wiley, New York, 1970.

KALWANI, M. U. AND A. J. SILK, "Structure of Repeat Buying for New Packaged Goods," *J. Marketing Res.*, 17 (1980), 316–322.

KARLIN, S. AND H. M. TAYLOR, *A First Course in Stochastic Processes*, Academic Press, New York, 1975.

LUKE, Y. L., *Algorithms for the Computation of Mathematical Functions*, Academic Press, New York, 1977.

MARITZ, J. S., *Empirical Bayes Methods*, Methuen, London, 1970.

MEYER, P. L., *Introductory Probability and Statistical Applications*, Addison-Wesley, Reading, Mass., 1970.

MORRIS, C. N., "Parametric Empirical Bayes Inference: Theory and Applications," *J. Amer. Statist. Assoc.*, 78 (1983), 47–55.

MORRISON, D. G., "Analysis of Consumer Purchase Data: A Bayesian Approach," *Industrial Management Rev.*, 9 (1968), 31–40.

———, "On Linearly Increasing Mean Residual Lifetimes," *J. Appl. Probab.*, 15 (1978), 617–620.

———, R. D. H. CHEN, S. L. KARPIS AND K. E. A. BRITNEY, "Modelling Retail Customer Behavior at Merrill Lynch," *Marketing Sci.*, 1 (1982), 123–141.

——— AND D. C. SCHMITTLEIN, "Predicting Future Random Events Based on Past Performance," *Management Sci.*, 27 (1981), 1006–1023.

SCHMITTLEIN, D. C. AND D. G. MORRISON, "The Median Residual Lifetime: A Characterization Theorem and an Application," *Oper. Res.*, 29 (1981), 392–399.

——— AND ———, "Prediction of Future Random Events with the Condensed Negative Binomial Distribution," *J. Amer. Statist. Assoc.*, 78 (1983), 449–456.

TRUSSELL, J. AND T. RICHARDS, "Correcting for Unmeasured Heterogeneity in Hazard Models Using the Heckman-Singer Procedure," in N. Tuma (Ed.) *Sociological Methodology 1985*, Jossey-Bass, San Francisco, Cal., 1985.