

# New Perspectives on Customer “Death” Using a Generalization of the Pareto/NBD Model

Kinshuk Jerath

Tepper School of Business, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, kinshuk@cmu.edu

Peter S. Fader

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

Bruce G. S. Hardie

London Business School, London NW1 4SA, United Kingdom, bhardie@london.edu

Several researchers have proposed models of buyer behavior in noncontractual settings that assume that customers are “alive” for some period of time and then become permanently inactive. The best-known such model is the Pareto/NBD, which assumes that customer attrition (dropout or “death”) can occur at any point in calendar time. A recent alternative model, the BG/NBD, assumes that customer attrition follows a Bernoulli “coin-flipping” process that occurs in “transaction time” (i.e., after every purchase occasion). Although the modification results in a model that is much easier to implement, it means that heavy buyers have more opportunities to “die.”

In this paper, we develop a model with a discrete-time dropout process tied to calendar time. Specifically, we assume that every customer periodically “flips a coin” to determine whether she “drops out” or continues as a customer. For the component of purchasing while alive, we maintain the assumptions of the Pareto/NBD and BG/NBD models. This *periodic death opportunity* (PDO) model allows us to take a closer look at how assumptions about customer death influence model fit and various metrics typically used by managers to characterize a cohort of customers. When the time period after which each customer makes her dropout decision (which we call *period length*) is very small, we show analytically that the PDO model reduces to the Pareto/NBD. When the period length is longer than the calibration period, the dropout process is “shut off,” and the PDO model collapses to the negative binomial distribution (NBD) model. By systematically varying the period length between these limits, we can explore the full spectrum of models between the “continuous-time-death” Pareto/NBD and the naïve “no-death” NBD.

In covering this spectrum, the PDO model performs at least as well as either of these models; our empirical analysis demonstrates the superior performance of the PDO model on two data sets. We also show that the different models provide significantly different estimates of both purchasing-related and death-related metrics for both data sets, and these differences can be quite dramatic for the death-related metrics. As more researchers and managers make managerial judgments that directly relate to the death process, we assert that the model employed to generate these metrics should be chosen carefully.

*Key words:* customer-base analysis; Pareto/NBD; BG/NBD; customer attrition

*History:* Received: February 9, 2011; accepted: March 16, 2011; Steve Shugan served as the acting editor-in-chief for this article. Published online in *Articles in Advance* June 6, 2011.

## 1. Introduction

As marketing researchers become more sophisticated in building models of customer behavior, they begin to scrutinize, test, and improve upon underlying assumptions of their models that were originally taken for granted. Within the domain of noncontractual customer-firm relationships, the assumptions made about the timing and nature of customer “death” (i.e., unobserved and unobservable dropout) are ripe for such improvements. For years, the gold standard for such models has been the Pareto/NBD (Schmittlein et al. 1987), which was the first to capture and exploit a customer death process in such a setting.

Applications of the model (e.g., Fader et al. 2005b, Reinartz and Kumar 2000, Schmittlein and Peterson 1994) have utilized the Pareto/NBD dropout process (namely, an exponential timing process with gamma-distributed heterogeneity across customers) without questioning it or testing alternative mechanisms.

The first paper to raise such questions was Fader et al. (2005a), which replaced the continuous-time gamma-exponential process operating in “calendar time” with a beta-geometric process operating in “transaction time” (i.e., the customer can only die immediately after a transaction). The resulting model, called the BG/NBD, was viewed as a “quick and

easy" alternative to the Pareto/NBD, because it offers a much more straightforward parameter estimation process with no substantial loss in the model's fit and forecasting capabilities. Beyond these computational benefits and aggregate indicators of overall model performance, however, not much attention was paid to the death process itself.

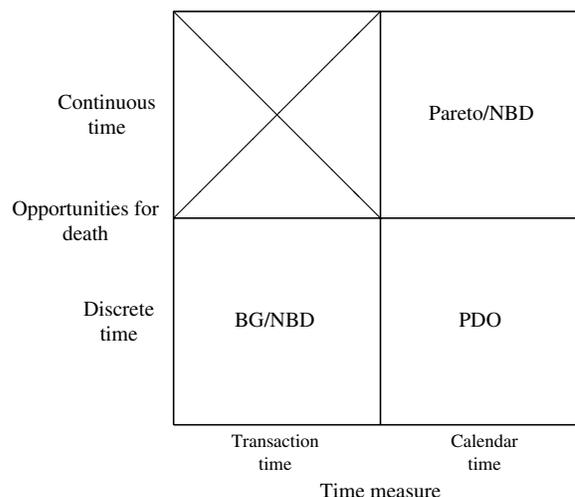
In this paper we propose and carefully investigate a new process for customer death in the noncontractual setting. The resulting model is a generalization of the Pareto/NBD and offers new insights about the death process. This new framework, called the *periodic death opportunity* (PDO) model, assumes that customers act in accordance with the discrete "coin-flipping" story associated with the geometric process, but these "coin flips" arise at periodic intervals in calendar time instead of transaction time (as in the BG/NBD model).

The differences in the assumed customer death process associated with these "buy-till-you-die" models of buyer behavior in noncontractual settings are illustrated in Figure 1. The first dimension on which they differ is whether the customer can "die" at any point in time (continuous) or only at discrete points in time. A second dimension appears when death occurs at discrete points in time: whether these points are discrete points in calendar time or are on a different timescale defined by the point in time at which transactions occurred. In all cases, it is assumed that opportunities for transactions (while alive) occur in continuous time; this is implicit in the use of the NBD to characterize buying behavior while alive.<sup>1</sup>

The PDO model allows us to take a closer look at how assumptions about customer death influence model fit as well as managerial inferences and diagnostics that emerge from the model. By varying the length of the time interval after which each customer makes her dropout decision, which we call the period length, we effectively vary the customer dropout dynamics. When the period length is very large, the dropout component is "shut off," and our model becomes the standard "no-death" negative binomial distribution (NBD) model (Ehrenberg 1959, Morrison and Schmittlein 1988). At the other extreme, when the period length tends to zero, our discrete-time dropout process converges into the continuous-time dropout process of the Pareto/NBD; we prove this convergence analytically. For intermediate values of period

<sup>1</sup> When the opportunities for transactions occur in discrete time (e.g., the NBD is replaced by the beta-binomial (BB) model), the "Continuous time" row in Figure 1 collapses into the "Discrete time" row (as time is discretized), and the distinction between "Calendar time" and "Transaction time" disappears (because they become one and the same timescale). Thus all three models collapse to the BG/BB model (Fader et al. 2010) when transactions can only occur at discrete points in time.

**Figure 1** Classifying Differences in Assumed Customer Death Processes



length, we can explore the full continuum of models that lie between these two extremes.

We find in our empirical analysis that the PDO model works better than both the NBD and the Pareto/NBD models for intermediate values of the period length, both in the calibration samples and longitudinal holdout periods. Hence, the PDO model acts as a unifying framework between the widely used Pareto/NBD and NBD models. We also find that the PDO model produces different estimates of some key managerial quantities compared to the alternative models. For instance, for both the data sets that we analyze, the PDO model predicts that the average customer in the cohort purchases slower while alive but lives longer, compared to the Pareto/NBD model.

We also compare the PDO model to the BG/NBD model and find that customer-death-related inferences are vastly different for the BG/NBD model. Recent research has focused on using death estimates obtained from buy-till-you-die models to augment managerial decisions regarding customer retention programs (Gopinath et al. 2009; Reinartz and Kumar 2000, 2003; Schweidel and Knox 2010; Wübben and Wangenheim 2008). Our results show that different models can lead to very different conclusions about customer death. In light of this, we assert that the model employed to generate these metrics should be chosen carefully, and the PDO model offers a rich framework for this analysis.

In the next section, we formally develop the PDO model. In §3, we carry out an empirical analysis in which the performance of the proposed model is compared to that of the Pareto/NBD, BG/NBD, and NBD models on two data sets—one using transactions from an online retailer of music CDs and the other using transactions from a grocery store. In §4, we extend the basic model to allow for heterogeneity in the period

length parameter and estimate this new model on both data sets. Finally, we conclude with a recap of the model, a brief discussion of its limitations, and some suggestions for related future research opportunities.

## 2. Model Development

The PDO model is based on the following six assumptions.

**ASSUMPTION 1.** *A customer's relationship with a specific firm can be characterized as first being alive for some period of time, then becoming permanently inactive (dead).*

**ASSUMPTION 2.** *While alive, the number of transactions made by a customer follows a Poisson process with mean  $\lambda$ . (This is equivalent to assuming that the inter-purchase times are independent and identically distributed exponential with rate  $\lambda$ .)*

**ASSUMPTION 3.** *Heterogeneity in transaction rates across customers follows a gamma distribution with shape parameter  $r$  and scale parameter  $\alpha$ :*

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

**ASSUMPTION 4.** *Let the random variable  $\Omega$  denote the unobserved time at which the customer dies. We model the death process by assuming that every  $\tau$  units of time (where time starts at zero), the customer can drop out with probability  $\theta$ . (This implies that the customer can drop out at  $\tau, 2\tau, \dots, \lfloor t/\tau \rfloor \tau$  in the interval  $(0, t]$ , where  $\lfloor \cdot \rfloor$  denotes the "floor" function.) Therefore, the probability that the customer has died by time  $t$  is*

$$P(\Omega \leq t | \theta, \tau) = 1 - (1 - \theta)^{\lfloor t/\tau \rfloor},$$

and the mean lifetime of the customer is  $E(\Omega | \theta, \tau) = \tau/\theta$ . We refer to  $\tau$  as the period length parameter and assume that it is the same for all customers. (We relax this assumption in §4.)

**ASSUMPTION 5.** *Heterogeneity in  $\theta$  follows a beta distribution with probability density function (pdf)*

$$f(\theta | a, b) = \frac{\theta^{a-1}(1 - \theta)^{b-1}}{B(a, b)}. \tag{2}$$

**ASSUMPTION 6.** *The transaction rate  $\lambda$  and the dropout probability  $\theta$  vary independently across customers.*

Note that the first three assumptions are identical to the corresponding assumptions of the Pareto/NBD model; the difference lies in the assumptions regarding the nature of the death process.

It follows from Assumptions 4 and 5 that the mean lifetime of a randomly chosen customer is

$$E(\Omega | a, b, \tau) = \tau \left( \frac{a + b - 1}{a - 1} \right), \tag{3}$$

and the probability that she died by  $t$  is

$$P(\Omega \leq t | a, b, \tau) = 1 - \frac{B(a, b + \lfloor t/\tau \rfloor)}{B(a, b)}. \tag{4}$$

The Pareto/NBD assumes that individual lifetimes follow an exponential distribution (in place of Assumption 4) and that heterogeneity in the underlying death rate follows a gamma distribution with shape parameter  $s$  and scale parameter  $\beta$  (in place of Assumption 5). This implies that, for a randomly chosen customer,

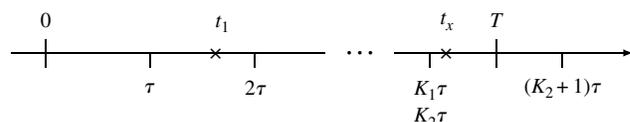
$$E(\Omega | s, \beta) = \frac{\beta}{s - 1}, \quad \text{and} \tag{5}$$

$$P(\Omega \leq t | s, \beta) = 1 - \left( \frac{\beta}{\beta + t} \right)^s. \tag{6}$$

On the face of it, these two models for the underlying death process seem quite different: in the PDO model a customer can die only at fixed points in time, whereas in the Pareto/NBD model a customer can die at any point in time. However, one can see that as  $\tau$  (in the PDO model) becomes smaller and smaller, the points in time when a customer can die come closer and closer. Extending this argument, as we let  $\tau$  approach zero, the customer can die at any point of time. The geometric "discrete" process then becomes an exponential "continuous" process. (We could think of the customers as continually flipping their coins to decide whether to drop out or not.) We can begin to see how the PDO model nests the Pareto/NBD as a special case.

Consider a customer who made  $x$  transactions in the interval  $(0, T]$ , with the transactions occurring at  $t_1, t_2, \dots, t_x$ ; by definition,  $t_x = 0$  when  $x = 0$ . We define  $K_1 = \lfloor t_x/\tau \rfloor$  and  $K_2 = \lfloor T/\tau \rfloor$ . Since  $t_x$  is the time of the last recorded transaction, the customer must have been alive at time  $K_1\tau$ , which implies that the customer definitely survived the first  $K_1$  opportunities to die. Because the customer was observed until time  $T$ ,  $K_2$  is the total number of opportunities the customer had to die (out of which she surely did not die in the first  $K_1$  opportunities). The meanings of, and relationships between,  $K_1$  and  $K_2$  will become clearer in the derivation below.

When  $K_1 = K_2$ , we have

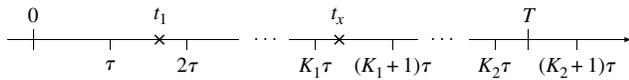


The fact that a purchase occurred at  $t_x$  implies the customer must have been alive at  $K_1\tau$ , which occurs with probability  $(1 - \theta)^{K_1}$ . Since  $K_1 = K_2$ , the customer

must still be alive at  $T$ . Given the model assumptions, the likelihood function for this case is

$$\begin{aligned} L(\lambda, \theta, \tau | t_1, \dots, t_x, T) &= \lambda e^{-\lambda t_1} \lambda e^{-\lambda(t_2-t_1)} \dots \lambda e^{-\lambda(t_x-t_{x-1})} e^{-\lambda(T-t_x)} (1-\theta)^{K_2} \\ &= \lambda^x e^{-\lambda T} (1-\theta)^{K_2}. \end{aligned}$$

When  $K_1 < K_2$ , we have



As before, the fact that a purchase occurred at  $t_x$  implies the customer must have been alive at  $K_1\tau$ . There are, however, a number of possible explanations for the lack of purchasing in the remaining interval  $(t_x, T]$ :

- The customer died at  $(K_1 + 1)\tau$ , having made no purchase in the interval  $(t_x, (K_1 + 1)\tau]$ , with likelihood function

$$\lambda^x e^{-\lambda(K_1+1)\tau} \theta (1-\theta)^{K_1}.$$

- The customer died at  $(K_1 + 2)\tau$ , having made no purchase in the interval  $(t_x, (K_1 + 2)\tau]$ , with likelihood function

$$\lambda^x e^{-\lambda(K_1+2)\tau} \theta (1-\theta)^{K_1+1}.$$

- The customer died at  $(K_1 + 3)\tau, \dots$

- The customer died at  $K_2\tau$ , having made no purchase in the interval  $(t_x, K_2\tau]$ , with likelihood function

$$\lambda^x e^{-\lambda K_2\tau} \theta (1-\theta)^{K_2-1}.$$

- The customer did not die at  $K_2\tau$  and made no purchase in the interval  $(t_x, T]$ , with likelihood function

$$\lambda^x e^{-\lambda T} (1-\theta)^{K_2}.$$

Note that in both cases ( $K_1 = K_2, K_1 < K_2$ ), information on when each of the  $x$  transactions occurred is not required; we can replace  $t_1, \dots, t_x, T$  with  $(x, t_x, T)$ . In other words,  $t_x$  and  $x$  are sufficient summaries of a customer's transaction history. (Using direct marketing terminology,  $t_x$  is recency and  $x$  is frequency.<sup>2</sup>)

Combining these two cases, we see that the individual-level likelihood function is

$$\begin{aligned} L(\lambda, \theta, \tau | x, t_x, T) &= \lambda^x e^{-\lambda T} (1-\theta)^{\lceil T/\tau \rceil} + \delta_{\lceil T/\tau \rceil > \lceil t_x/\tau \rceil} \\ &\cdot \sum_{j=1}^{\lceil T/\tau \rceil - \lceil t_x/\tau \rceil} \lambda^x e^{-\lambda(\lceil t_x/\tau \rceil + j)\tau} \theta (1-\theta)^{\lceil t_x/\tau \rceil + j - 1}. \end{aligned} \quad (7)$$

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

Taking the expectation of (7) over the distributions of  $\lambda$  and  $\theta$ , (1) and (2), results in the following expression for the likelihood function for a randomly chosen customer with purchase history  $(x, t_x, T)$ :

$$\begin{aligned} L(r, \alpha, a, b, \tau | x, t_x, T) &= \int_0^1 \int_0^\infty L(\lambda, \theta, \tau | x, t_x, T) f(\lambda | r, \alpha) f(\theta | a, b) d\lambda d\theta \\ &= \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)} \left[ \left( \frac{1}{\alpha+T} \right)^{r+x} \frac{B(a, b + \lceil T/\tau \rceil)}{B(a, b)} + \delta_{\lceil T/\tau \rceil > \lceil t_x/\tau \rceil} \right. \\ &\quad \cdot \left. \sum_{j=1}^{\lceil T/\tau \rceil - \lceil t_x/\tau \rceil} \left\{ \left( \frac{1}{\alpha + (\lceil t_x/\tau \rceil + j)\tau} \right)^{r+x} \cdot \frac{B(a+1, b + \lceil t_x/\tau \rceil + j - 1)}{B(a, b)} \right\} \right]. \end{aligned} \quad (8)$$

The five model parameters ( $r, \alpha, a, b$ , and  $\tau$ ) can be estimated via the method of maximum likelihood in the following manner. Suppose we have a sample of  $I$  customers, where customer  $i$  had  $x_i$  transactions in the interval  $(0, T_i]$ , with the last transaction occurring at  $t_{x_i}$ . The sample log-likelihood function is given by

$$LL(r, \alpha, a, b, \tau) = \sum_{i=1}^I \ln[L(r, \alpha, a, b, \tau | x_i, t_{x_i}, T_i)]. \quad (9)$$

We find the maximum of this function using a line search for  $\tau$  coupled with standard optimization methods for the other four parameters.

In Appendix A, we show that as  $\tau \rightarrow 0$ , (8) becomes the likelihood function associated with the Pareto/NBD model. As  $\tau \rightarrow \infty$ ,  $\lceil T/\tau \rceil = 0$  and  $\delta_{\lceil T/\tau \rceil > \lceil t_x/\tau \rceil} = 0$ , in which case (8) collapses to

$$\frac{\Gamma(r+x)\alpha^r}{\Gamma(r)} \left( \frac{1}{\alpha+T} \right)^{r+x},$$

which is simply the timing-model analog of the basic NBD model (Gupta and Morrison 1991). (Strictly speaking, the PDO model collapses to the NBD whenever  $\tau > \max\{T_i, i = 1, \dots, I\}$ .)

Following Schmittlein et al. (1987), three quantities of managerial interest in a customer-base analysis exercise are as follows.

- The expected number of transactions in a time interval of length  $t$  is given by

$$\begin{aligned} E(X(t) | r, \alpha, a, b, \tau) &= \frac{rt}{\alpha} \frac{B(a, b + \lceil t/\tau \rceil)}{B(a, b)} + \frac{r\tau}{\alpha} \sum_{j=1}^{\lceil t/\tau \rceil} j \frac{B(a+1, b+j-1)}{B(a, b)}. \end{aligned} \quad (10)$$

(This quantity is central to computing the expected transaction volume for the whole customer base over time.)

- The probability that a customer with observed behavior  $(x, t_x, T)$  is still alive at time  $T$  is given by

$$P(\Omega > T | r, \alpha, a, b, \tau, x, t_x, T) = \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)} \left( \frac{1}{\alpha+T} \right)^{r+x} \cdot \frac{B(a, b + \lfloor T/\tau \rfloor)}{B(a, b)} / L(r, \alpha, a, b, \tau | x, t_x, T). \quad (11)$$

- The expected number of transactions in the interval  $(T, T+t]$  for a customer with observed behavior  $(x, t_x, T)$  is given by

$$E(X(T, T+t) | r, \alpha, a, b, \tau, x, t_x, T) = \frac{1}{L(r, \alpha, a, b, \tau | x, t_x, T)} \times \frac{\Gamma(r+x+1)\alpha^r}{\Gamma(r)} \left( \frac{1}{\alpha+T} \right)^{r+x+1} \cdot \left\{ \frac{B(a, b + \lfloor (T+t)/\tau \rfloor)}{B(a, b)} t + \sum_{j=1}^{\lfloor (T+t)/\tau \rfloor - \lfloor T/\tau \rfloor} \frac{B(a+1, b + \lfloor T/\tau \rfloor + j - 1)}{B(a, b)} \cdot [(\lfloor T/\tau \rfloor + j)\tau - T] \right\}. \quad (12)$$

(See Appendix B for the associated derivations.)

### 3. Empirical Analysis

We now examine the performance of the PDO model using the CDNOW data set used in Fader et al. (2005a, b) and a grocery purchase data set used in Batislam et al. (2007) (henceforth, the Grocery data set).

#### 3.1. The CDNOW Data Set

This data set tracks 2,357 individuals who made their first-ever purchases at the CDNOW website in the first 12 weeks of 1997 and records their repeat purchasing through June 1998. The first 39 weeks of data are used for model calibration; the remaining 39 weeks of data are used as longitudinal holdout for model validation. Fitting the Pareto/NBD model to these data yields a log-likelihood of  $-9,595.0$ ; fitting the NBD model yields a log-likelihood of  $-9,763.7$ . Clearly, the Pareto/NBD model does much better than the NBD model while using only two extra parameters (likelihood ratio test  $\chi_2^2 = 337.4$ ,  $p < 0.001$ ). Our focus, however, is on the performance of the PDO model. Is its fit bounded between these Pareto/NBD and NBD limits, or does it provide a superior fit to the data?

Varying  $\tau$  from 0.01 weeks to 40 weeks (in increments of 0.001), we find the maximum likelihood estimates of the remaining four model parameters by maximizing the log-likelihood function given in (9).

The corresponding values of the log-likelihood function are plotted in Figure 2. This figure shows that the log-likelihood function is discontinuous in  $\tau$  at multiple values of  $\tau$ .<sup>3</sup> These discontinuities occur as a result of the discrete nature of the floor functions in (8), which have  $\tau$  as an argument. At certain values of  $\tau$ , a very small increase in  $\tau$  causes a large discrete change in the values of the floor functions (i.e., for a very small increase in  $\tau$ , the value of the floor function can decrease by one for several cohort members), which leads to a discontinuity in the sample log-likelihood function in (9). (More specifically, this occurs when  $t_x$  is a multiple of  $\tau$ .) Despite the presence of discontinuities in the sample log-likelihood function, the following patterns are clearly discernible.

When  $\tau$  is very small ( $\tau = 0.01$  weeks,  $LL = -9,595.3$ ), the fit of the PDO model is almost exactly the same as that of the Pareto/NBD. As  $\tau$  increases from 0.01, the log-likelihood increases from the Pareto/NBD limit to a maximum value of  $-9,585.6$  at  $\tau = 3.001$ , then starts declining toward the much lower value associated with the NBD.<sup>4</sup> As soon as  $\tau > 6$  weeks, the fit of the PDO model is strictly worse than that of the Pareto/NBD. When  $\tau$  is large ( $>39$  weeks in this empirical setting), the PDO model yields the same log-likelihood as that of the NBD model.

The PDO model (with  $\tau = 3.001$ ) provides a significant improvement in calibration-period model fit over the Pareto/NBD (likelihood ratio test  $\chi_1^2 = 18.8$ ,  $p < 0.001$ ). The parameter estimates of the best-fitting PDO model and the Pareto/NBD model are reported in Table 1.<sup>5,6</sup>

To illustrate more clearly the relationship between the Pareto/NBD, the NBD, and the PDO models with various values of  $\tau$ , we present the parameter estimates and other summary statistics for these models

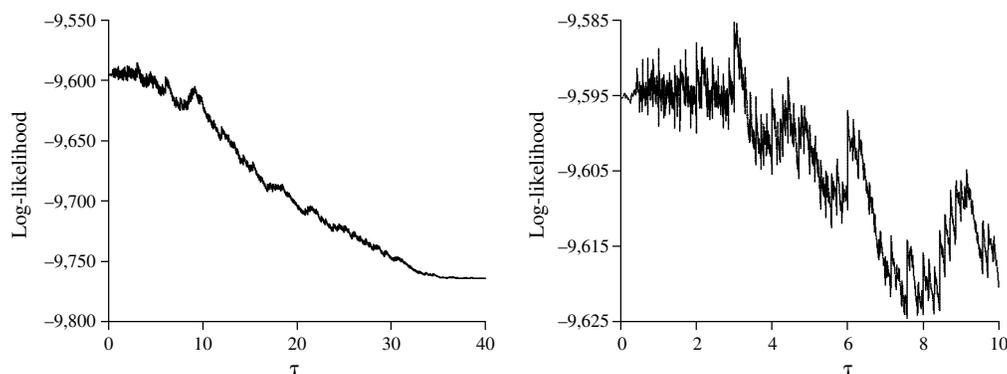
<sup>3</sup> We thank Albert Bemmaor and Nicolas Glady for bringing this point to our attention.

<sup>4</sup> We could use an adaptive line search that switches to a finer incremental in the region of the maximum value of  $\tau$ . For example, if we switch to an increment of  $1e-06$  in the region of  $\tau = 3.001$ , we find that the optimal value of  $\tau$  is 3.000001 with an associated log-likelihood of  $-9,585.3$ . (There is no change in the other parameter estimates to two decimal places and no change in any derived quantities of managerial interest, such as forecasts of repeat sales.) This implies the "truly" optimal value of  $\tau$  is  $3 + \epsilon$ , where  $\epsilon \rightarrow 0^+$ . For the sake of simplicity, we stay with the nonadaptive line search with an increment of 0.001 for the analyses presented in this paper.

<sup>5</sup> Note that the log-likelihood function for the PDO model is discontinuous around  $\tau = 3.001$  (where it attains its maximum). The standard errors that we report here are obtained by local smoothing of this function in the neighborhood of  $\tau = 3.001$  (Barnett 1966, Daniels 1961). Specifically, we report the values provided by the `fmincon` function in Matlab, which uses quadratic local smoothing. Other methods, such as bootstrap or jackknife, may also be used to obtain standard errors.

<sup>6</sup> The NBD parameter estimates are  $\hat{r} = 0.38$  and  $\hat{a} = 12.07$ .

**Figure 2** Plot of the PDO Model Log-Likelihood for the CDNOW Data Set as a Function of the Period Length Parameter  $\tau$



Note. The right-hand plot presents a “zoomed-in” view for small values of  $\tau$ .

in Table 2. Note how the parameters of the PDO model tend toward those of the Pareto/NBD model as  $\tau \rightarrow 0$  ( $a \rightarrow s, b\tau \rightarrow \beta$ ). This is in accordance with the pattern predicted theoretically in Appendix A.

Beyond the raw parameter estimates, we also see an interesting range of values for two summary statistics of the underlying behavioral characteristics—namely, the mean transaction rate and the median lifetime.

**Table 1** Parameter Estimates and Standard Errors for the PDO and Pareto/NBD Models for the CDNOW Data Set

	PDO		Pareto/NBD	
	Estimate	Std. error	Estimate	Std. error
$r$	0.52	0.008	0.55	0.026
$\alpha$	10.40	0.008	10.58	0.632
$a$	0.43	0.015		
$b$	2.61	0.002		
$\tau$	3.001	0.002		
$s$			0.61	0.025
$\beta$			11.67	0.307
LL	-9,585.6		-9,595.0	

**Table 2** Estimation Results for the CDNOW Data Set for the Pareto/NBD Model ( $\tau = 0$ ), the NBD Model ( $\tau \rightarrow \infty$ ), and the PDO Model (for Various Values of  $\tau$ )

$\tau$	Transaction		Death		LL	$E(\Lambda)$	Median( $\Omega$ )
	$r$	$\alpha$	$a^a$	$b^a$			
(0)	0.55	10.58	0.61	11.67	-9,595.0	0.052	25.0
0.010	0.55	10.58	0.61	1,167.41	-9,595.3	0.052	25.0
0.100	0.55	10.57	0.60	115.29	-9,595.5	0.052	25.0
1.001	0.54	10.50	0.56	10.90	-9,589.7	0.051	26.5
2.001	0.52	10.16	0.53	5.20	-9,588.3	0.051	26.5
3.001	0.52	10.40	0.43	2.61	-9,585.6	0.050	27.8
4.001	0.50	10.45	0.42	2.02	-9,594.3	0.048	29.8
5.001	0.48	11.26	0.46	2.03	-9,599.2	0.043	30.8
10.001	0.45	10.49	0.26	0.62	-9,620.0	0.043	45.0
20.001	0.40	11.18	0.63	1.51	-9,703.5	0.036	53.8
( $\infty$ )	0.39	12.07	—	—	-9,763.7	0.032	—

<sup>a</sup> $s$  and  $\beta$  in the case of the Pareto/NBD model ( $\tau = 0$ ).

First, comparing the two extreme models, we expect that under the Pareto/NBD, we would have shorter lifetimes and therefore higher average transaction rates when compared with the NBD. Pursuing this logic further, it follows that increases in  $\tau$  would be associated with declining average transaction rates and increasing median lifetimes. What is not known, however, is how quickly these statistics will change with  $\tau$  and whether the rates of change are similar or different when we compare them. The data in Table 2 confirm the expected directional changes for these two statistics as a function of  $\tau$ , but they also reveal some differences. As  $\tau$  rises from 0 (i.e., Pareto/NBD model) to 3.001, the mean transaction rate falls by about 5%, but the median lifetime rises by about 10%. These differences are not dramatic, but they suggest that the estimated death rates, rather than the mean transaction rates, are more sensitive to changes in  $\tau$ .

Because the reduction in transaction rate is smaller than the increase in lifetime, the best-fitting PDO model generates a slightly higher estimate of total repeat sales over time than that produced by the Pareto/NBD model. But is this good or bad? As  $\tau$  rises from 0 (i.e., Pareto/NBD model) to 3.001, the mean transaction rate falls by about 5%, but the median lifetime rises by about 10%. To examine this we create total repeat sales forecasts for each of the specifications reported in Table 2. In Table 3 we report the mean absolute percentage error (MAPE) numbers for both cumulative total repeat sales and weekly total repeat sales over weeks 40–78.

Looking at the cumulative MAPE numbers, we see that all of the PDO models with a finite period length forecast the cumulative sales trajectory extremely well; it is hard to discern meaningful differences in a plot of these curves. In contrast, when we look at errors on an incremental (week-by-week) basis, there are greater deviations (as would be expected). Overall, however, there is strong empirical support for

**Table 3** Measures of Model Forecasting Performance for the Pareto/NBD Model ( $\tau = 0$ ), the NBD Model ( $\tau \rightarrow \infty$ ) and the PDO Model (for Various Values of  $\tau$ )

$\tau$	MAPE (Weeks 40–78)	
	Cumulative (%)	Weekly (%)
(0)	1.35	20.89
0.010	1.35	20.89
0.100	1.34	20.53
1.001	1.14	20.13
2.001	1.09	19.81
3.001	0.85	19.06
4.001	0.85	19.06
5.001	0.88	19.33
10.001	0.70	19.18
20.001	0.92	19.25
( $\infty$ )	10.37	36.22

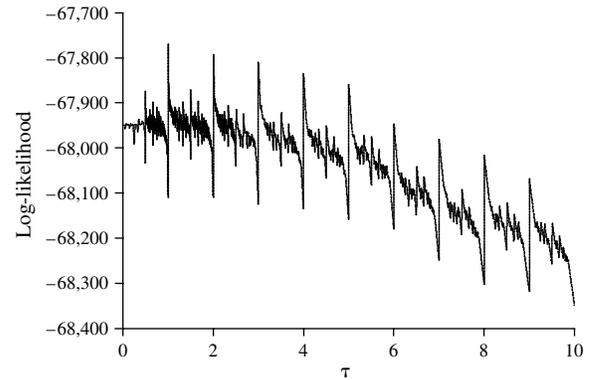
the performance of the PDO model (particularly with  $\tau = 3.001$ ) as a worthy alternative to the Pareto/NBD.

Do these results prove that customers are actually “flipping their coins” every three weeks? Of course not; however, they do suggest that there is room for improvement in modeling the death process beyond the starting point established by the Pareto/NBD, and they provide some reasonable evidence to support the general idea of the discrete-time “story” being told here. The consistency of these results and their superiority to a very strong benchmark (represented by the Pareto/NBD) are hard to deny.

### 3.2. The Grocery Data Set

Batislam et al. (2007) used a data set covering 5,479 individuals at a Turkish grocery store who made their first-ever purchases between August 2001 and October 2001, recording their repeat purchasing through April 2003. For every individual, we have data on the recency and frequency of repeat purchasing and the length of time for which this individual was observed. To make our study consistent with that of Batislam et al. (2007), we use the first 78 weeks for calibration and the last 13 weeks as longitudinal holdout for model validation.

Fitting the Pareto/NBD model to these data yields a log-likelihood of  $-67,925.8$ , and fitting the NBD model yields  $-71,000.5$ . Fitting the PDO models by varying  $\tau$  from 0.01 weeks to 80 weeks, we confirm the pattern we observed for the CDNOW data set—when  $\tau$  is very small (i.e.,  $\tau = 0.01$  weeks), the fit of the PDO model is identical to that of the Pareto/NBD; when  $\tau$  is large ( $>78$  weeks in this empirical setting), the PDO model yields the same log-likelihood as that of the NBD model. The values of the log-likelihood function as  $\tau$  varies from 0.01 to 10 weeks are plotted in Figure 3. (This is the equivalent of the “zoomed-in” right-hand plot in Figure 2.) Note that, as in the

**Figure 3** Plot of the PDO Model Log-Likelihood for the Grocery Data Set as a Function of the Period Length Parameter  $\tau$ 

case of the CDNOW data set, the log-likelihood function for the Grocery data set is discontinuous at multiple values of  $\tau$ . The maximum log-likelihood value is achieved when  $\tau = 1.001$  and is  $-67,757.3$ . Compared with the Pareto/NBD model, this is a highly significant improvement in model fit at the cost of just one additional parameter (likelihood ratio test  $\chi_1^2 = 337$ ,  $p < 0.001$ ). In Table 4, we present the parameter estimates of the best-fitting PDO model and the Pareto/NBD model.<sup>7</sup>

As expected, we also observe exactly the same patterns as in Table 2 for the parameter estimates, mean underlying transaction rates, and median lifetimes for the different PDO models as  $\tau$  varies, and we observe the same qualitative relationships between these models and the Pareto/NBD model. For instance, for the best-fitting PDO model, the mean underlying transaction rate is 0.11 per week and the median lifetime is 45.2 weeks; for the Pareto/NBD model, the mean underlying transaction rate is the same (0.11 per week) and the median lifetime is smaller (41.8 weeks). Furthermore, for the 13-week holdout period also, we observe similar patterns as in Table 3. For the PDO ( $\tau = 1.001$ ) model, the cumulative MAPE is 0.9% and the weekly MAPE is 3.6%; for the Pareto/NBD model, the cumulative MAPE is 0.5% and the weekly MAPE is 3.7%.

To summarize, the best-fitting PDO model performs significantly better than the Pareto/NBD model in terms of in-sample fit for the Grocery data set, and we observe similar relationships between the Pareto/NBD, NBD, and various PDO models in this data set as we did in the CDNOW data set.

### 3.3. Comparison with the BG/NBD Model

A natural question that arises is, what about the original BG/NBD model (with “coin flips” tied to transactions instead of occurring periodically in calendar

<sup>7</sup> The NBD parameter estimates are  $\hat{\tau} = 0.31$  and  $\hat{\alpha} = 4.86$ .

**Table 4** Parameter Estimates and Standard Errors for the PDO and Pareto/NBD Models for the Grocery Data Set

	PDO		Pareto/NBD	
	Estimate	Std. error	Estimate	Std. error
$r$	0.46	0.001	0.48	0.001
$\alpha$	4.38	0.019	4.38	0.014
$a$	0.62	0.004		
$b$	22.19	0.011		
$\tau$	1.001	0.000		
$s$			0.57	0.002
$\beta$			17.64	0.031
$LL$	-67,757.3		-67,925.8	

time)? The answer is quite interesting. In terms of in-sample fit, the BG/NBD performs at least as well as all of the PDO models for the CDNOW data set; its log-likelihood value of  $-9,582.4$ , as reported in Fader et al. (2005a), is slightly better than that of the PDO model with  $\tau = 3.001$  (and therefore substantially better than that of the regular Pareto/NBD). However, for the Grocery data set, the BG/NBD model ( $LL = -68,007.0$ ) significantly underperforms both the PDO model and the Pareto/NBD model.<sup>8</sup> In terms of out-of-sample performance on the summary statistics shown in Table 3, the BG/NBD consistently performs slightly worse than the PDO model—for the CDNOW data set, the BG/NBD model’s cumulative MAPE is 2.6% and weekly MAPE is 19.4% (both slightly worse), and for the Grocery data set, the BG/NBD model’s cumulative MAPE is 8.8% (much worse) and weekly MAPE is 4.0% (slightly worse). Overall, these fit statistics offer mixed evidence about the superiority of the PDO model vis-à-vis the BG/NBD, and none of it is particularly dramatic.

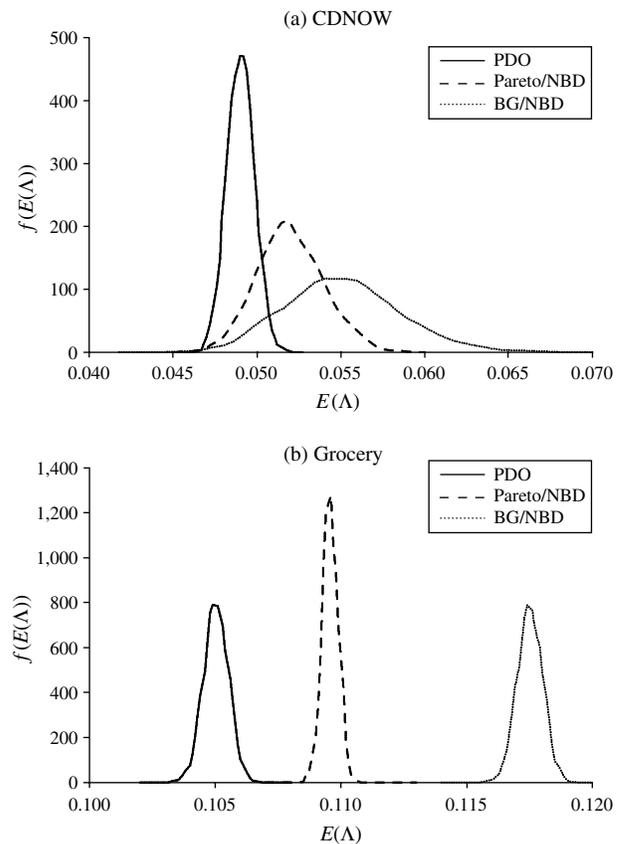
Further insight into the three models can be obtained by comparing inferences we may draw concerning underlying buyer behavior. In particular, we compare the estimates for the mean transaction rate and the median lifetime with uncertainty in the maximum-likelihood (ML) estimates taken into account.<sup>9</sup>

Figure 4 plots the estimates for the mean transaction rate with uncertainty in the ML estimates taken into account. For both data sets, there are notable differences in the mean purchasing rates; a randomly

<sup>8</sup> The BG/NBD model parameter estimates, along with the standard errors in parentheses, are  $\hat{r} = 0.24$  (0.013),  $\hat{\alpha} = 4.41$  (0.366),  $\hat{a} = 0.79$  (0.189), and  $\hat{b} = 2.43$  (0.695) for the CDNOW data set; and  $\hat{r} = 0.28$  (0.004),  $\hat{\alpha} = 2.34$  (0.008),  $\hat{a} = 0.40$  (0.022), and  $\hat{b} = 2.09$  (0.178) for the Grocery data set.

<sup>9</sup> To account for uncertainty in the estimates, we take 1,000 draws of the relevant parameters from a multivariate normal distribution, where the means are the ML estimates of the relevant parameters and the covariance matrix is given by the inverse of the Hessian (to account for the correlations between parameter estimates).

**Figure 4** Expected (Latent) Transaction Rate for a Randomly Chosen Customer for the CDNOW and Grocery Data Sets, Accounting for Uncertainty in the ML Estimates, for the Best-Fitting PDO, Pareto/NBD and BG/NBD Models



chosen customer purchases slower (while alive) under the PDO model compared to the Pareto/NBD model and purchases slower (while alive) under the Pareto/NBD model compared to the BG/NBD model. (We note that the estimates for these quantities are sharper for the Grocery data set than the CDNOW data set because in the former we have data for almost twice the number of individuals compared with the latter.)

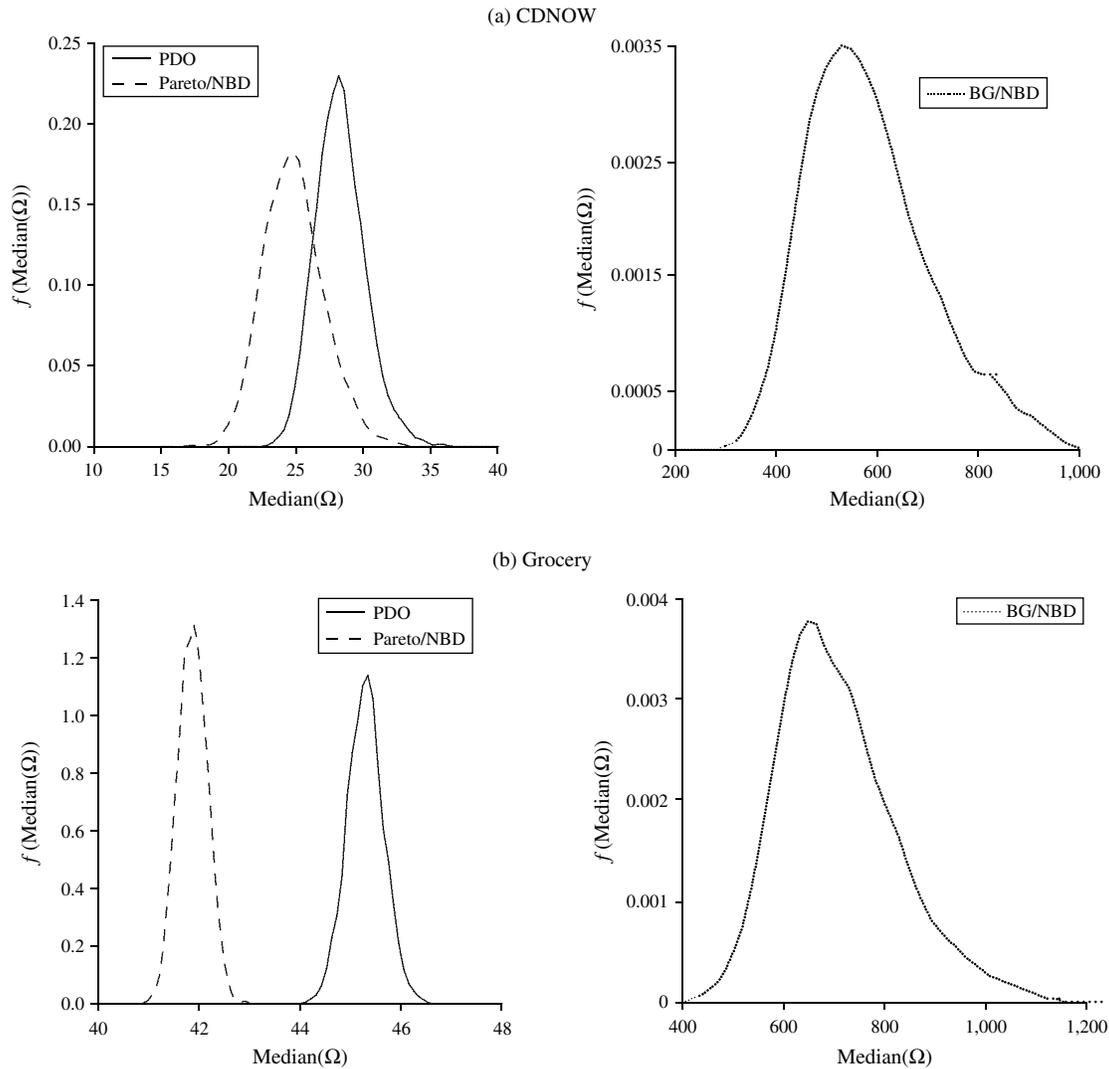
Noting that the BG/NBD model probability of a randomly chosen customer having died by  $t$  is given by

$$P(\Omega \leq t) = 1 - \left( \frac{\alpha}{\alpha + t} \right)^r {}_2F_1 \left( r, b; a + b; \frac{t}{\alpha + t} \right), \quad (13)$$

where  ${}_2F_1(\cdot)$  is the Gaussian hypergeometric function, we can compute the median lifetime for all three models with uncertainty in the ML estimates taken into account—see Figure 5 for the associated plots.

In both empirical settings, there is a slight (but systematic) difference in the median lifetimes as implied by the PDO and Pareto/NBD models, with the PDO model predicting that a randomly chosen customer in the cohort lives longer; however, the results for the

**Figure 5** Median (Latent) Lifetime for a Randomly Chosen Customer for the CDNOW and Grocery Data Sets, Accounting for Uncertainty in the ML Estimates, for the Best-Fitting PDO, Pareto/NBD, and BG/NBD Models



BG/NBD model cast a dubious light on its validity—even when its overall fit is superior. Specifically, we note that, for both data sets, the median lifetimes as implied by the BG/NBD are so vastly different than those of the two former models that they cannot be plotted on the same scale.

Compared with the Pareto/NBD model, the PDO model predicts (in these empirical settings) that a randomly chosen customer in the cohort lives longer but purchases slower while alive. Under the BG/NBD model, a randomly chosen customer not only lives much longer but also purchases at a slightly higher rate. How do we reconcile this with the finding that the BG/NBD underestimates cumulative repeat sales when compared to the Pareto/NBD model (Batislam et al. 2007, Fader et al. 2005a)? In contrast to the cumulative death probability for the PDO and Pareto/NBD models, (4) and (6), respectively, the cumulative death probability for the BG/NBD model,

(13), is a function of the parameters of the distribution that characterizes heterogeneity in buying rate while alive. This follows from the fact that death can only follow a purchase. Although the mean of the gamma heterogeneity distribution is highest for the BG/NBD model for both data sets, it is also the most heterogeneous (as indicated by the coefficient of variation, which equals  $1/\sqrt{r}$ ). This pool of customers with a very low buying rate means that their opportunities for death occur very infrequently, which results in very long (unobserved) lifetimes.

These questionable results for the BG/NBD highlight the practical implications of the death process associated with the class of buy-till-you-die models; assessing measures of model fit is not sufficient to judge the suitability of a model for inferential purposes. As more researchers and managers make managerial judgments that directly relate to the death process (Gopinath et al. 2009; Reinartz

and Kumar 2000, 2003; Schweidel and Knox 2010; Wübben and Wangenheim 2008), the theoretical benefits of the PDO framework become much more tangible. The BG/NBD has no direct connection to the Pareto/NBD, and there is no way to equate the parameters across these different specifications. These conceptual benefits help demonstrate the value of the PDO framework, especially for researchers with a focal interest in the nature of the death process.

#### 4. Heterogeneity in $\tau$

In the previous sections, we have treated the parameter  $\tau$  as homogeneous for all customers; in reality, however, the nature of the period length is likely to vary across them. In this section, we allow for heterogeneity in  $\tau$ . Specifically, we assume that  $\tau$  varies across customers according to a gamma distribution with pdf

$$f(\tau | m, \rho) = \frac{\rho^m \tau^{m-1} e^{-\rho\tau}}{\Gamma(m)}. \quad (14)$$

(Using other heterogeneity distributions, such as a log-normal distribution or a normal distribution truncated below at zero, yields similar results.)

Under this specification, the individual-level process is the same as before, and the likelihood function for a randomly chosen customer is obtained by taking the expectation of (8) over the distribution of  $\tau$ , i.e., evaluating the integral

$$\int_0^\infty L(r, \alpha, a, b, \tau) f(\tau | m, \rho) d\tau.$$

As this does not have an algebraic solution, we turn to Markov chain Monte Carlo (MCMC) methods for estimating this heterogeneous PDO (henceforth HPDO) model; see Appendix C for details.

We estimate the HPDO model on both data sets. Our inferences are based on 50,000 iterations, following a burn-in of 20,000 iterations. The resulting

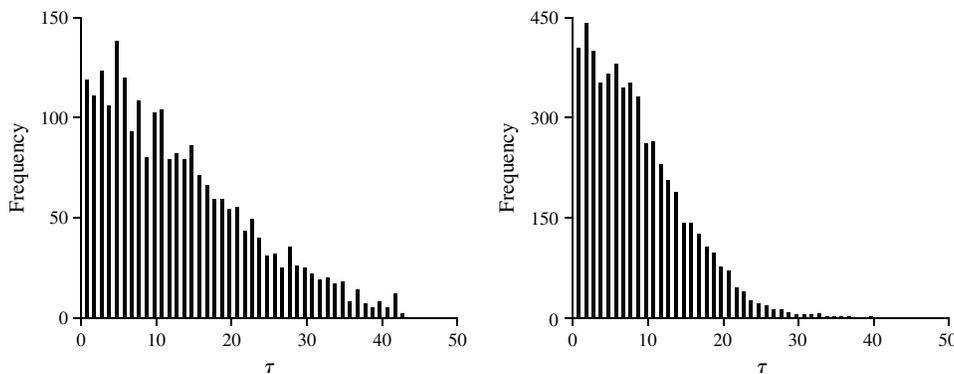
**Table 5** Parameter Estimates for the HPDO Model for the CDNOW and Grocery Data Sets

	CDNOW		Grocery	
	Posterior mean	95% interval	Posterior mean	95% interval
$r$	0.41	[0.408, 0.413]	0.35	[0.349, 0.351]
$\alpha$	11.74	[11.647, 11.833]	4.55	[4.529, 4.571]
$a$	0.20	[0.194, 0.206]	0.54	[0.535, 0.545]
$b$	2.75	[2.648, 2.852]	10.54	[10.354, 10.726]
$m$	1.47	[1.463, 1.478]	1.42	[1.416, 1.424]
$\rho$	0.12	[0.118, 0.123]	0.21	[0.207, 0.214]

parameter estimates after the Markov chains converge are presented in Table 5. (The associated log-marginal density numbers, calculated using the harmonic mean of the likelihoods across iterations (see Newton and Raftery 1994), are  $-8,577.1$  and  $-63,746.2$  for the CDNOW and Grocery data sets, respectively.)

Figure 6 shows the distribution of  $\tau$  for both data sets. Clearly, there is significant heterogeneity in  $\tau$  in both data sets. Note that the expected value of  $\tau$  is larger for the CDNOW data set compared with the Grocery data set; this is in line with the value of  $\tau$  for the best-fitting PDO model (i.e., without heterogeneity in  $\tau$ ) being larger for the CDNOW data set compared to the Grocery data set. However, although the HPDO model offers new insights into the dropout process and a large improvement in in-sample fit, it performs slightly worse in terms of out-of-sample predictions than the PDO models without heterogeneity in  $\tau$ . For the CDNOW data set, for weeks 40–78, the cumulative MAPE is 1.92% and the weekly MAPE is 20.55%. For the Grocery data set, for weeks 78–91, the cumulative MAPE is 1.5% and the weekly MAPE is 8.1%. Furthermore, using simulated data for various “HPDO worlds,” we found that the HPDO model is unable to do a satisfactory job of recovering the true individual-level period lengths, which we conjecture is because the individual-level likelihood function has floor functions that make it discontinuous in  $\tau$ . Similar simulations for the basic PDO model

**Figure 6** Histogram of Posterior Draws of  $\tau$  Across Customers from One Iteration After Convergence for the CDNOW Data Set (Left) and the Grocery Data Set (Right)



covering a wide variety of purchasing characteristics in the cohorts show that, in contrast to the HPDO model, it is able to recover its original parameters with good accuracy in spite of the discontinuity issues that we discussed in §3. (In the case of the basic PDO model, we are able to conduct a line search to find the optimal value of  $\tau$  for the cohort.) For these reasons, the simpler PDO model may be the preferred alternative. Nevertheless, we believe that the HPDO model is a worthwhile extension to consider and might offer value to future researchers, particularly if they want to test hypotheses about how the nature of the dropout process varies across customers (perhaps related to customer characteristics or contextual factors).

## 5. Discussion and Conclusions

We have proposed the PDO model as a new way to better understand and capture the death process associated with buy-till-you-die type models that are frequently used to model customer buying behavior in noncontractual settings. We demonstrated (both analytically and empirically) that the PDO model nests both the Pareto/NBD and the traditional no-death NBD as special cases, and we found strong evidence that customers behave as if their death process is somewhere in between these two extremes. Because the PDO model has a more flexible death process, it widens the scope of what we can infer about the nature of the customer dropout process when compared to the Pareto/NBD model. Because the purchase process is estimated jointly with the death process, we can also expect to see differences in inferences for the quantities related to the purchase process. In accordance with these insights, we found that the PDO model can indeed offer different estimates for some important characteristics of purchase and death dynamics in a cohort—for instance, for both the data sets that we analyzed, we found that the average customer lives longer but purchases slower in the PDO model than in the Pareto/NBD model.

Although this paper questions (and improves upon) one aspect of the original Pareto/NBD paradigm, it does not necessarily mean that the Pareto/NBD itself is obsolete and should be avoided. We continue to encourage using the Pareto/NBD model when the manager's primary goal is forecasting purchases, as opposed to a principal focus on the death process, per se. Although in our data sets the best-fitting PDO models (and the HPDO model) offer consistent and significant improvements in the calibration sample, the improvements in predictions for the holdout period are not especially dramatic. On the other hand, when inferences about customer dropout are central to the manager's goals, the PDO framework deserves her attention.

This paper has focused on the underlying structure of the death process in the buy-till-you-die models of buyer behavior in noncontractual settings, as illustrated in Figure 1. All the models assume that an individual customer's unobserved (and unobservable) lifetime is characterized by a memoryless process (i.e., an exponential or geometric distribution) and that buying behavior while alive can be characterized by the Poisson distribution coupled with gamma heterogeneity (i.e., the NBD). Staying within the buy-till-you-die paradigm, there are obvious extensions. We can replace the memoryless lifetime distribution (associated with any of the three cells in Figure 1) with one that exhibits some form of duration dependence. For instance, if we find that the Pareto/NBD is underpredicting repeat sales or overpredicting the number of zero repeat buyers, or both, we could replace the exponential lifetime distribution with a Weibull (or some other distribution with a more flexible hazard function that has the ability to "slow down" the death process). We can also relax the assumption of a memoryless interpurchase distribution associated with the NBD. For example, the NBD can be replaced by the CNBD (Chatfield and Goodhardt 1973, Schmittlein and Morrison 1983), as in Platzer's (2008) CBG/CNBD-k model and the work of Schweidel and Knox (2010).

The emergence of simulation-based estimation methods for the Pareto/NBD model (Abe 2009, Ma and Liu 2007, Singh et al. 2009) means that it is now much easier for the researcher to tinker with the basic "building blocks" of the model, i.e., the ability to utilize other probability distributions instead of those used by the early researchers in this area. Although this kind of flexibility seems like a welcome addition to the modeler's toolkit, it may create an overly complex model that lacks a meaningful "as-if" story. We feel that a more reflective approach to extending the model, such as stepping back and questioning the implicit assumption that "death can occur at any time," may result in better performance (particularly in a holdout period) as well as improved diagnostics (such as the nature of the death process).

A logical extension of our hierarchical Bayesian estimation procedure for the HPDO model would be the use of covariates to help explain how the death process varies across customers. Several researchers have explored such extensions to the Pareto/NBD model using time-invariant covariates (e.g., Abe 2009, Fader and Hardie 2007) and time-varying covariates (e.g., Schweidel and Knox 2010). This could help managers manage customer attrition in an effective manner. These issues, however, go beyond the PDO model proposed here (and may not offer substantial benefits to justify their additional complexity). Before rushing ahead with these potential improvements, we

encourage researchers to contemplate the basic PDO model and take advantage of its desirable theoretical properties.

### Acknowledgments

The authors thank Emine Batislam, Meltem Denizel, and Alpay Filiztekin for making the Grocery data set available and Albert Bemmaor, Nicolas Glady, and Marcel Goić for helpful discussions. The first author acknowledges the support of the i-Lab at Carnegie Mellon University, the second author acknowledges the support of the Wharton Customer Analytics Initiative, and the third author acknowledges the support of the London Business School Centre for Marketing.

### Appendix A. The Pareto/NBD as a Limiting Case

In this section we show that the PDO model likelihood function approaches that of the Pareto/NBD model as  $\tau \rightarrow 0$ .

The individual-level likelihood function for the Pareto/NBD model is (Fader and Hardie 2005):

$$L(\lambda, \mu | x, t_x, T) = \frac{\lambda^x \mu}{\lambda + \mu} e^{-(\lambda + \mu)t_x} + \frac{\lambda^{x+1}}{\lambda + \mu} e^{-(\lambda + \mu)T}.$$

Assuming heterogeneity in  $\lambda$  and  $\mu$  is distributed gamma with parameters  $(r, \alpha)$  and  $(s, \beta)$ , respectively, it follows that the likelihood function for a randomly chosen customer is found by solving

$$\begin{aligned} L(r, \alpha, s, \beta | x, t_x, T) &= \int_0^\infty \int_0^\infty \left( \frac{\lambda^x \mu}{\lambda + \mu} e^{-(\lambda + \mu)t_x} \right. \\ &\quad \left. + \frac{\lambda^{x+1}}{\lambda + \mu} e^{-(\lambda + \mu)T} \right) f(\lambda | r, \alpha) f(\mu | s, \beta) d\mu d\lambda. \end{aligned} \quad (A1)$$

Our proof is based on showing that

$$\lim_{\tau \rightarrow 0} \int_0^1 \int_0^\infty L(\lambda, \theta, \tau | x, t_x, T) f(\lambda | r, \alpha) f(\theta | a, b) d\lambda d\theta$$

is identical to (A1) (i.e.,  $\lim_{\tau \rightarrow 0} L(r, \alpha, a, b, \tau) \equiv L(r, \alpha, s, \beta)$ ).

This proof will make use of the following results:

$$\lim_{\tau \rightarrow 0} (1 - \mu\tau)^{T/\tau} = e^{-\mu T}, \quad (A2)$$

$$\lim_{\tau \rightarrow 0} \frac{e^{\lambda\tau} - (1 - \mu\tau)}{\tau} = \lambda + \mu, \quad (A3)$$

$$\lim_{\tau \rightarrow 0} \frac{\Gamma(s + \beta/\tau)}{\Gamma(\beta/\tau)} \tau^s = \beta^s. \quad (A4)$$

(We note that (A2) is a standard result, (A3) results from the application of L'Hôpital's rule, and (A4) follows from Abramowitz and Stegun 1972, Equation 6.1.46.)

Noting that

$$\begin{aligned} &\sum_{j=1}^{\lfloor T/\tau \rfloor - \lfloor t_x/\tau \rfloor} \lambda^x e^{-\lambda(\lfloor t_x/\tau \rfloor + j)\tau} \theta(1 - \theta)^{\lfloor t_x/\tau \rfloor + j - 1} \\ &= \lambda^x e^{-\lambda(\lfloor t_x/\tau \rfloor + 1)\tau} \theta(1 - \theta)^{\lfloor t_x/\tau \rfloor} \sum_{l=0}^{\lfloor T/\tau \rfloor - \lfloor t_x/\tau \rfloor - 1} [e^{-\lambda\tau} (1 - \theta)]^l \\ &= \lambda^x e^{-\lambda(\lfloor t_x/\tau \rfloor + 1)\tau} \theta(1 - \theta)^{\lfloor t_x/\tau \rfloor} \frac{1 - [e^{-\lambda\tau} (1 - \theta)]^{\lfloor T/\tau \rfloor - \lfloor t_x/\tau \rfloor}}{1 - [e^{-\lambda\tau} (1 - \theta)]} \\ &= \frac{\lambda^x e^{-\lambda(\lfloor t_x/\tau \rfloor)\tau} \theta(1 - \theta)^{\lfloor t_x/\tau \rfloor}}{e^{\lambda\tau} - (1 - \theta)} - \frac{\lambda^x e^{-\lambda(\lfloor T/\tau \rfloor)\tau} \theta(1 - \theta)^{\lfloor T/\tau \rfloor}}{e^{\lambda\tau} - (1 - \theta)}, \end{aligned}$$

we can rewrite the individual-level likelihood function (7) as

$$\begin{aligned} L(\lambda, \theta, \tau | x, t_x, T) &= \lambda^x e^{-\lambda T} (1 - \theta)^{\lfloor T/\tau \rfloor} + \delta_{\lfloor T/\tau \rfloor > \lfloor t_x/\tau \rfloor} \left\{ \frac{\lambda^x e^{-\lambda \lfloor t_x/\tau \rfloor \tau} \theta(1 - \theta)^{\lfloor t_x/\tau \rfloor}}{e^{\lambda\tau} - (1 - \theta)} \right. \\ &\quad \left. - \frac{\lambda^x e^{-\lambda \lfloor T/\tau \rfloor \tau} \theta(1 - \theta)^{\lfloor T/\tau \rfloor}}{e^{\lambda\tau} - (1 - \theta)} \right\}. \end{aligned}$$

Let  $\theta = \mu\tau$ , which implies  $d\theta = \tau d\mu$ . (When  $\theta = 0, \mu = 0$ ; similarly, when  $\theta = 1, \mu = 1/\tau$ .) Also let  $a = s$  and  $b = \beta/\tau$ . It follows that PDO likelihood function for a randomly chosen customer,  $L(r, \alpha, a, b, \tau)$ , can be written as

$$\begin{aligned} L(r, \alpha, s, \beta, \tau) &= \int_0^{1/\tau} \int_0^\infty \left\{ L(\lambda, \mu\tau, \tau | x, t_x, T) f(\lambda | r, \alpha) \right. \\ &\quad \left. \times (\mu\tau)^{s-1} (1 - \mu\tau)^{\beta/\tau - 1} \frac{\Gamma(s + \beta/\tau)}{\Gamma(s)\Gamma(\beta/\tau)} \right\} d\lambda (\tau d\mu) \\ &= \mathcal{A}_1 + \delta_{\lfloor T/\tau \rfloor > \lfloor t_x/\tau \rfloor} (\mathcal{A}_2 - \mathcal{A}_3), \end{aligned} \quad (A5)$$

where

$$\begin{aligned} \mathcal{A}_1 &= \int_0^{1/\tau} \int_0^\infty \lambda^x e^{-\lambda T} \tau^s \mu^{s-1} (1 - \mu\tau)^{\lfloor T/\tau \rfloor + \beta/\tau - 1} \\ &\quad \cdot \frac{\Gamma(s + \beta/\tau)}{\Gamma(s)\Gamma(\beta/\tau)} f(\lambda | r, \alpha) d\lambda d\mu, \\ \mathcal{A}_2 &= \int_0^{1/\tau} \int_0^\infty \frac{\lambda^x e^{-\lambda \lfloor t_x/\tau \rfloor \tau} \tau^s \mu^s (1 - \mu\tau)^{\lfloor t_x/\tau \rfloor + \beta/\tau - 1}}{e^{\lambda\tau} - (1 - \mu\tau)} \\ &\quad \cdot \frac{\Gamma(s + \beta/\tau)}{\Gamma(s)\Gamma(\beta/\tau)} f(\lambda | r, \alpha) d\lambda d\mu, \\ \mathcal{A}_3 &= \int_0^{1/\tau} \int_0^\infty \frac{\lambda^x e^{-\lambda \lfloor T/\tau \rfloor \tau} \tau^s \mu^s (1 - \mu\tau)^{\lfloor T/\tau \rfloor + \beta/\tau - 1}}{e^{\lambda\tau} - (1 - \mu\tau)} \\ &\quad \cdot \frac{\Gamma(s + \beta/\tau)}{\Gamma(s)\Gamma(\beta/\tau)} f(\lambda | r, \alpha) d\lambda d\mu. \end{aligned}$$

(Note that the only difference between  $\mathcal{A}_2$  and  $\mathcal{A}_3$  is  $t_x$  versus  $T$ .)

We now take the limit of (A5) as  $\tau \rightarrow 0$ . Noting that  $\delta_{\lfloor T/\tau \rfloor > \lfloor t_x/\tau \rfloor} = 1$  as  $\tau \rightarrow 0$ , we have

$$\lim_{\tau \rightarrow 0} L(r, \alpha, s, \beta, \tau) = \lim_{\tau \rightarrow 0} \mathcal{A}_1 + \lim_{\tau \rightarrow 0} \mathcal{A}_2 - \lim_{\tau \rightarrow 0} \mathcal{A}_3. \quad (A6)$$

Noting  $\lfloor t_x/\tau \rfloor \simeq t_x/\tau$  and  $\lfloor T/\tau \rfloor \simeq T/\tau$  as  $\tau \rightarrow 0$ , and using (A2)–(A4),

$$\begin{aligned} \lim_{\tau \rightarrow 0} \mathcal{A}_1 &= \lim_{\tau \rightarrow 0} \int_0^\infty \int_0^{1/\tau} \lambda^x e^{-\lambda T} \tau^s \mu^{s-1} (1 - \mu\tau)^{\lfloor T/\tau \rfloor + \beta/\tau - 1} \\ &\quad \cdot \frac{\Gamma(s + \beta/\tau)}{\Gamma(s)\Gamma(\beta/\tau)} f(\lambda | r, \alpha) d\mu d\lambda \\ &= \int_0^\infty \int_0^1 \lim_{\tau \rightarrow 0} \left\{ \lambda^x e^{-\lambda T} \tau^s \mu^{s-1} (1 - \mu\tau)^{\lfloor T/\tau \rfloor + \beta/\tau - 1} \right. \\ &\quad \left. \cdot \frac{\Gamma(s + \beta/\tau)}{\Gamma(s)\Gamma(\beta/\tau)} f(\lambda | r, \alpha) \right\} d\mu d\lambda \\ &= \int_0^\infty \int_0^\infty \lambda^x e^{-\lambda T} \frac{\beta^s \mu^{s-1} e^{-\mu(\beta+T)}}{\Gamma(s)} f(\lambda | r, \alpha) d\mu d\lambda \\ &= \int_0^\infty \int_0^\infty \lambda^x e^{-(\lambda + \mu)T} f(\lambda | r, \alpha) f(\mu | s, \beta) d\mu d\lambda, \end{aligned} \quad (A7)$$

and

$$\begin{aligned} \lim_{\tau \rightarrow 0} \mathcal{A}_2 &= \lim_{\tau \rightarrow 0} \int_0^\infty \int_0^{1/\tau} \frac{\lambda^x e^{-\lambda \lfloor t_x/\tau \rfloor \tau} \tau^s \mu^s (1-\mu\tau)^{\lfloor t_x/\tau \rfloor + \beta/\tau - 1}}{e^{\lambda\tau} - (1-\mu\tau)} \\ &\quad \cdot \frac{\Gamma(s+\beta/\tau)}{\Gamma(s)\Gamma(\beta/\tau)} f(\lambda | r, \alpha) d\mu d\lambda \\ &= \int_0^\infty \int_0^{\lim_{\tau \rightarrow 0} 1/\tau} \lim_{\tau \rightarrow 0} \left\{ \frac{\lambda^x e^{-\lambda \lfloor t_x/\tau \rfloor \tau} \tau^s \mu^s (1-\mu\tau)^{\lfloor t_x/\tau \rfloor + \beta/\tau - 1}}{e^{\lambda\tau} - (1-\mu\tau)} \right. \\ &\quad \left. \cdot \frac{\Gamma(s+\beta/\tau)}{\Gamma(s)\Gamma(\beta/\tau)} f(\lambda | r, \alpha) \right\} d\mu d\lambda \\ &= \int_0^\infty \int_0^\infty \frac{\lambda^x e^{-\lambda t_x} \beta^s \mu^s e^{-\mu(\beta+t_x)}}{\lambda + \mu} \frac{1}{\Gamma(s)} f(\lambda | r, \alpha) d\mu d\lambda \\ &= \int_0^\infty \int_0^\infty \frac{\lambda^x \mu e^{-(\lambda+\mu)t_x}}{\lambda + \mu} f(\lambda | r, \alpha) f(\mu | s, \beta) d\mu d\lambda. \quad (\text{A8}) \end{aligned}$$

It follows that

$$\lim_{\tau \rightarrow 0} \mathcal{A}_3 = \int_0^\infty \int_0^\infty \frac{\lambda^x \mu e^{-(\lambda+\mu)T}}{\lambda + \mu} f(\lambda | r, \alpha) f(\mu | s, \beta) d\mu d\lambda. \quad (\text{A9})$$

Substituting (A7)–(A9) in (A6) and simplifying gives us

$$\begin{aligned} \lim_{\tau \rightarrow 0} L(r, \alpha, s, \beta, \tau) &= \int_0^\infty \int_0^\infty \left( \frac{\lambda^x \mu}{\lambda + \mu} e^{-(\lambda+\mu)t_x} + \frac{\lambda^{x+1}}{\lambda + \mu} e^{-(\lambda+\mu)T} \right) \\ &\quad \cdot f(\lambda | r, \alpha) f(\mu | s, \beta) d\mu d\lambda, \end{aligned}$$

which is exactly the integral for the Pareto/NBD likelihood function, (A1). Q.E.D.

This proof establishes the equivalence between the likelihoods of observing the same data under the PDO model when  $\tau \rightarrow 0$  and the Pareto/NBD. We use this fact somewhat liberally to “prove” that the two models are equivalent under this special condition ( $\tau \rightarrow 0$ ). (Using a similar procedure and the same substitutions, it is easy to show that all the expressions for the PDO model—e.g., (10)–(12)—are identical to those of the Pareto/NBD as  $\tau \rightarrow 0$ .)

## Appendix B. Derivations of Key Results

**Derivation of (10).** Let the random variable  $X(t)$  denote the number of transactions occurring in the interval  $(0, t]$ . Conditional on  $\lambda$ , it follows from the assumption of Poisson purchasing that  $E[X(t)]$  is simply  $\lambda t$  if the customer is active at  $t$ ,  $\lambda\tau$  if the customer dies at  $\tau$ ,  $2\lambda\tau$  if the customer dies at  $2\tau$ ,  $3\lambda\tau$  if the customer dies at  $3\tau$ , ..., and  $\lfloor t/\tau \rfloor \lambda\tau$  if the customer dies at  $\lfloor t/\tau \rfloor \tau$ . Multiplying these quantities by the probability that the customer dies at  $\tau, 2\tau, \dots$ , gives us

$$E(X(t) | \lambda, \theta, \tau) = \lambda t (1 - \theta)^{\lfloor t/\tau \rfloor} + \lambda\tau \sum_{j=1}^{\lfloor t/\tau \rfloor} \theta (1 - \theta)^{j-1}.$$

Taking the expectation of this over the distributions of  $\lambda$  and  $\theta$ , (1) and (2), gives us the expression in (10).

**Derivation of (11).** The probability that a customer with purchase history  $(x, t_x, T)$  is alive at time  $T$  is simply the probability that she was alive at  $K_2\tau$ . Referring back to our derivation of the individual-level likelihood function, (7), the application of the Bayes’ theorem gives us

$$P(\Omega > T | \lambda, \theta, \tau, x, t_x, T) = \frac{\lambda^x e^{-\lambda T} (1 - \theta)^{\lfloor T/\tau \rfloor}}{L(\lambda, \theta, \tau | x, t_x, T)}. \quad (\text{B1})$$

(We note that if  $\lfloor t_x/\tau \rfloor = \lfloor T/\tau \rfloor$  (i.e.,  $K_1 = K_2$ ),  $P(\Omega > T | \lambda, \theta, \tau, x, t_x, T) = 1$ .)

By Bayes’ theorem, the joint posterior distribution of  $\lambda$  and  $\theta$  is given by

$$\begin{aligned} f(\lambda, \theta | r, \alpha, a, b, \tau, x, t_x, T) &= \frac{L(\lambda, \theta, \tau | x, t_x, T) f(\lambda | r, \alpha) f(\theta | a, b)}{L(r, \alpha, a, b, \tau | x, t_x, T)}. \quad (\text{B2}) \end{aligned}$$

Taking the expectation of (B1) over this joint posterior distribution gives us (11).

**Derivation of (12).** Let the random variable  $X(T, T + t)$  denote the number of purchases made in the interval  $(T, T + t]$ . We are interested in computing the conditional expectation  $E(X(T, T + t) | x, t_x, T)$ , the expected number of purchases in the interval  $(T, T + t]$  for a customer with purchase history  $(x, t_x, T)$ .

Let us assume the customer is alive at  $T$  (i.e.,  $\Omega > T$ ). Conditional on  $\lambda$ , it follows from the assumption of Poisson purchasing that the expected number of purchases in  $(T, T + t]$  is simply  $\lambda t$  if the customer is active at  $T + t$ ,  $\lambda[(\lfloor T/\tau \rfloor + 1)\tau - T]$  if the customer dies at  $(\lfloor T/\tau \rfloor + 1)\tau$ ,  $\lambda[(\lfloor T/\tau \rfloor + 2)\tau - T]$  if the customer dies at  $(\lfloor T/\tau \rfloor + 2)\tau, \dots$ , and  $\lambda[(\lfloor T + t \rfloor/\tau)\tau - T]$  if the customer dies at  $\lfloor (T + t)/\tau \rfloor \tau$ . Multiplying these quantities by the probability that the customer dies at  $(\lfloor T/\tau \rfloor + 1), (\lfloor T/\tau \rfloor + 2), \dots$ , gives us

$$\begin{aligned} E(X(T, T + t) | \lambda, \theta, \tau, \Omega > T) &= \lambda t (1 - \theta)^{\lfloor (T+t)/\tau \rfloor - \lfloor T/\tau \rfloor} \\ &\quad + \sum_{j=1}^{\lfloor (T+t)/\tau \rfloor - \lfloor T/\tau \rfloor} \lambda [(\lfloor T/\tau \rfloor + j)\tau - T] \theta (1 - \theta)^{j-1}. \quad (\text{B3}) \end{aligned}$$

Taking the expectation of the product of (B1) and (B3) over the joint posterior distribution of  $\lambda$  and  $\theta$ , (B2), gives us (12).

## Appendix C. MCMC Procedure for the HPDO Model

Customer  $i$  with period length  $\tau_i$  makes her dropout decision at  $\tau_i, 2\tau_i, 3\tau_i, \dots$ . Whether the customer dropped out or not and, if so, when she dropped out, is unobserved. To aid with the model estimation, we generate these unobservables using data augmentation (Tanner and Wong 1987). Specifically, we use the indicator variable  $Z_i$  (with the realization  $z_i$ ) to denote whether the customer died between  $t_{x_i}$  and  $T_i$ : if the customer is still alive at  $T_i$ ,  $Z_i = 1$ ; if the customer died at  $\psi_i \tau_i$ , where  $\psi_i \in \{\lfloor t_{x_i}/\tau_i \rfloor + 1, \dots, \lfloor T_i/\tau_i \rfloor\}$ ,  $Z_i = 0$ .

Recalling the logic of the derivation of (7), the likelihood function of customer  $i$  is given by  $\lambda_i^{x_i} e^{-\lambda_i T_i} (1 - \theta_i)^{\lfloor T_i/\tau_i \rfloor}$  if  $Z_i = 1$ , and  $\lambda_i^{x_i} e^{-\lambda_i \psi_i \tau_i} \theta_i (1 - \theta_i)^{\psi_i - 1}$  if  $Z_i = 0$  (and the customer died at  $\psi_i \tau_i$ ). Therefore,

$$\begin{aligned} L(\lambda_i, \theta_i, \tau_i | x_i, t_{x_i}, T_i, z_i, \psi_i) &= \lambda_i^{x_i} e^{-\lambda_i [z_i T_i + (1-z_i) \psi_i \tau_i]} \theta_i^{(1-z_i)} (1 - \theta_i)^{\lfloor z_i \lfloor T_i/\tau_i \rfloor + (1-z_i)(\psi_i - 1)}. \end{aligned}$$

The parameters of the heterogeneity distributions for  $\lambda$ ,  $\theta$ , and  $\tau$  specified earlier act as priors for  $\lambda_i$ ,  $\theta_i$ , and  $\tau_i$ , respectively. Hence,  $\lambda_i \sim \text{gamma}(r, \alpha)$ ,  $\theta_i \sim \text{beta}(a, b)$ , and  $\tau_i \sim \text{gamma}(m, \rho)$ . We derive expressions for the conditional densities of the relevant individual-level parameters in the following manner.

• The conditional posterior distribution of  $\lambda_i$  is proportional to  $L(\lambda_i, \theta_i, \tau_i | x_i, t_{x_i}, T_i, z_i, \psi_i) \times f(\lambda_i | r, \alpha)$ , which in turn is proportional to  $\lambda_i^{r+x_i-1} e^{-\lambda_i[\alpha+z_i T_i+(1-z_i)\psi_i \tau_i]}$ . Therefore,

$$\lambda_i | r, \alpha, \tau_i, x_i, T_i, z_i, \psi_i \sim \text{gamma}(r + x_i - 1, \alpha + z_i T_i + (1 - z_i)\psi_i \tau_i). \quad (C1)$$

• The conditional posterior distribution of  $\theta_i$  is proportional to  $L(\lambda_i, \theta_i, \tau_i | x_i, t_{x_i}, T_i, z_i, \psi_i) \times f(\theta_i | a, b)$ , which in turn is proportional to  $\theta_i^{a-z_i}(1-\theta_i)^{b-1+z_i[T_i/\tau_i]+(1-z_i)(\psi_i-1)}$ . Therefore,

$$\theta_i | a, b, \tau_i, T_i, z_i, \psi_i \sim \text{beta}(a + 1 - z_i, b + z_i \lfloor T_i/\tau_i \rfloor + (1 - z_i)(\psi_i - 1)). \quad (C2)$$

• The conditional posterior distribution of  $\tau_i$  is proportional to  $L(\lambda_i, \theta_i, \tau_i | x_i, t_{x_i}, T_i, z_i, \psi_i) \times f(\tau_i | m, \rho)$ , which gives us

$$f(\tau_i | m, \rho, \tau_i, T_i, z_i, \psi_i) \propto \tau_i^{m-1} e^{-[\rho\tau_i + \lambda_i(z_i T_i + (1-z_i)\psi_i \tau_i)](1-\theta_i)^{z_i \lfloor T_i/\tau_i \rfloor + (1-z_i)(\psi_i-1)}}. \quad (C3)$$

We can sample from this distribution using Metropolis-Hastings methods.

- Recalling (B1),

$$P(Z_i = 1 | \lambda_i, \theta_i, \tau_i, t_{x_i}, T_i) = \frac{\mathcal{B}_1}{\mathcal{B}_1 + \delta_{\lfloor T_i/\tau_i \rfloor > \lfloor t_{x_i}/\tau_i \rfloor} \mathcal{B}_2}, \quad (C4)$$

where

$$\mathcal{B}_1 = e^{-\lambda_i T_i} (1 - \theta_i)^{\lfloor T_i/\tau_i \rfloor}, \quad \text{and}$$

$$\mathcal{B}_2 = \sum_{j=1}^{\lfloor T_i/\tau_i \rfloor - \lfloor t_{x_i}/\tau_i \rfloor} e^{-\lambda_i (\lfloor t_{x_i}/\tau_i \rfloor + j)\tau_i} \theta_i (1 - \theta_i)^{\lfloor t_{x_i}/\tau_i \rfloor + j - 1}.$$

Note that if  $\lfloor T_i/\tau_i \rfloor = \lfloor t_{x_i}/\tau_i \rfloor$ , customer  $i$  is definitely alive at  $T_i$  because she did not get a chance to flip her death coin after her last purchase (which occurred at  $t_{x_i}$ ); as such,  $P(Z_i = 1 | \cdot) = 1$ .

• If  $Z_i = 0$ , then we also generate an integer  $\psi_i \in \{\lfloor t_{x_i}/\tau_i \rfloor + 1, \dots, \lfloor T_i/\tau_i \rfloor\}$ , where  $t_{x_i} < \psi_i \tau_i < T_i$  is the point in time at which customer  $i$  died. Assuming a (discrete) uniform prior on  $\psi_i$ ,

$$P(\Psi_i = \psi_i | \lambda_i, \theta_i, \tau_i, t_{x_i}, T_i) \propto e^{-\lambda_i \psi_i \tau_i} \theta_i (1 - \theta_i)^{\psi_i - 1}. \quad (C5)$$

The joint conditional posterior distributions of the population-level parameters are given by

$$f(r, \alpha | \boldsymbol{\lambda}) \propto \left\{ \prod_{i=1}^I f(\lambda_i | r, \alpha) \right\} f(r) f(\alpha), \quad (C6)$$

$$f(a, b | \boldsymbol{\theta}) \propto \left\{ \prod_{i=1}^I f(\theta_i | a, b) \right\} f(a) f(b), \quad (C7)$$

$$f(m, \rho | \boldsymbol{\tau}) \propto \left\{ \prod_{i=1}^I f(\tau_i | m, \rho) \right\} f(m) f(\rho), \quad (C8)$$

where  $\boldsymbol{\lambda}$ ,  $\boldsymbol{\theta}$ , and  $\boldsymbol{\tau}$  denote the current vectors of the individual-level parameters. We use fully diffuse hyperpriors for  $r$ ,  $\alpha$ ,  $a$ ,  $b$ ,  $m$ , and  $\rho$ . (Although these parameters are only allowed to take on positive values, in our case this is

not a problem because the two data sets we use have a large number of individuals in them; i.e., while sampling from the posterior distribution, the data likelihood strongly dominates the prior specification.) Metropolis-Hastings methods are used to sample the pairs  $(r, \alpha)$ ,  $(a, b)$ , and  $(m, \rho)$ .

The resulting MCMC procedure used in the estimation is as follows:

1. Set initial values for  $r, \alpha, a, b, m$ , and  $\rho$ , as well as for  $\lambda_i, \theta_i, \tau_i, z_i$ , and  $\psi_i, \forall i = 1, \dots, I$ .
2. Iterate until convergence:
  - (a) For each customer  $i$ , sample  $\lambda_i, \theta_i$ , and  $\tau_i$  using (C1), (C2), and (C3), respectively.
  - (b) For each customer  $i$ , generate  $z_i$  using (C4). If  $z_i = 0$ , generate  $\psi_i$  using (C5). (This is the data augmentation step.)
  - (c) Sample the pairs  $(r, \alpha)$ ,  $(a, b)$ , and  $(m, \rho)$  using (C6), (C7), and (C8), respectively.

## References

- Abe, M. 2009. "Counting your customers" one by one: A hierarchical Bayes extension to the Pareto/NBD model. *Marketing Sci.* 28(3) 541–553.
- Abramowitz, M., I. A. Stegun, eds. 1972. *Handbook of Mathematical Functions*. Dover Publications, New York.
- Barnett, V. D. 1966. Evaluation of the maximum-likelihood estimator where the likelihood equation has multiple roots. *Biometrika* 53(1/2) 151–165.
- Batıslam, E. P., M. Denizel, A. Filiztekin. 2007. Empirical validation and comparison of models for customer base analysis. *Internat. J. Res. Marketing* 24(3) 201–209.
- Chatfield, C., G. J. Goodhardt. 1973. A consumer purchasing model with Erlang inter-purchase times. *J. Amer. Statist. Assoc.* 68(344) 828–835.
- Daniels, H. E. 1961. The asymptotic efficiency of a maximum likelihood estimator. *Proc. 4th Berkeley Sympos. Math. Statist. Probab.*, Vol. 1. University of California Press, Berkeley, 151–163.
- Ehrenberg, A. S. C. 1959. The pattern of consumer purchases. *Appl. Statist.* 8(1) 26–41.
- Fader, P. S., B. G. S. Hardie. 2005. A note on deriving the Pareto/NBD model and related expressions. Retrieved April 10, 2010, <http://brucehardie.com/notes/009/>.
- Fader, P. S., B. G. S. Hardie. 2007. Incorporating time-invariant covariates into the Pareto/NBD and BG/NBD models. Retrieved April 10, 2010, <http://brucehardie.com/notes/019/>.
- Fader, P. S., B. G. S. Hardie, K. L. Lee. 2005a. "Counting your customers" the easy way: An alternative to the Pareto/NBD model. *Marketing Sci.* 24(2) 275–284.
- Fader, P. S., B. G. S. Hardie, K. L. Lee. 2005b. RFM and CLV: Using iso-value curves for customer base analysis. *J. Marketing Res.* 42(4) 415–430.
- Fader, P. S., B. G. S. Hardie, J. Shang. 2010. Customer-base analysis in a discrete-time noncontractual setting. *Marketing Sci.* 29(6) 1086–1108.
- Gopinath, S., R. Blattberg, E. Malthouse. 2009. Are revived customers as good as new? Working paper, Northwestern University, Evanston, IL. <http://ssrn.com/abstract=1356851>.
- Gupta, S., D. G. Morrison. 1991. Estimating heterogeneity in consumers' purchase rates. *Marketing Sci.* 10(3) 264–269.
- Ma, S.-H., J.-L. Liu. 2007. The MCMC approach for solving the Pareto/NBD model and possible extensions. *Proc. 3rd Internat. Conf. Natural Comput. (ICNC 2007)*, IEEE Computer Society, Washington, DC, 505–512.

- Morrison, D. G., D. C. Schmittlein. 1988. Generalizing the NBD model for customer purchases: What are the implications and is it worth the effort? *J. Bus. Econom. Statist.* 6(2) 145–159.
- Newton, M. A., A. E. Raftery. 1994. Approximate Bayesian inference with the weighted likelihood bootstrap. *J. Roy. Statist. Soc. Ser. B* 56(1) 3–48.
- Platzer, M. 2008. Stochastic models of noncontractual consumer relationships. Master's thesis, Vienna University of Economics and Business Administration, Vienna.
- Reinartz, W., J. V. Kumar. 2000. On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing. *J. Marketing* 64(4) 17–35.
- Reinartz, W. J., V. Kumar. 2003. The impact of customer relationship characteristics on profitable lifetime duration. *J. Marketing* 67(1) 77–99.
- Schmittlein, D. C., D. G. Morrison. 1983. Prediction of future random events with the condensed negative binomial distribution. *J. Amer. Statist. Assoc.* 78(382) 449–456.
- Schmittlein, D. C., R. A. Peterson. 1994. Customer base analysis: An industrial purchase process application. *Marketing Sci.* 13(1) 41–67.
- Schmittlein, D. C., D. G. Morrison, R. Colombo. 1987. Counting your customers: Who are they and what will they do next? *Management Sci.* 33(1) 1–24.
- Schweidel, D. A., G. Knox. 2010. Incorporating strategic direct marketing activity into "buy 'til you die" models. Working paper, University of Wisconsin–Madison, Madison. <http://ssrn.com/abstract=1670060>.
- Singh, S. S., S. Borle, D. C. Jain. 2009. A generalized framework for estimating customer lifetime value when customer lifetimes are not observed. *Quant. Marketing Econom.* 7(2) 181–205.
- Tanner, M. A., W. H. Wong. 1987. The calculation of posterior distributions by data augmentation. *J. Amer. Statist. Assoc.* 82(398) 528–540.
- Wübben, M., F. v. Wangenheim. 2008. Instant customer base analysis: Managerial heuristics often "get it right." *J. Marketing* 72(3) 82–93.