New Perspectives on Customer "Death" Using a Generalization of the Pareto/NBD Model

Kinshuk Jerath
Peter S. Fader
Bruce G. S. Hardie[†]

June 2007
Revised May 2009

[†]Kinshuk Jerath is Assistant Professor of Marketing at the Tepper School of Business, Carnegie Mellon University (email: kinshuk@cmu.edu). Peter S. Fader is the Frances and Pei-Yuan Chia Professor of Marketing at the Wharton School of the University of Pennsylvania (address: 749 Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104-6340; phone: 215.898.1132; email: faderp@wharton.upenn.edu; web: www.petefader.com). Bruce G. S. Hardie is Professor of Marketing, London Business School (email: bhardie@london.edu; web: www.brucehardie.com). The authors thank Emine Batislam, Meltem Denizel and Alpay Filiztekin for making the Grocery dataset available. The second author acknowledges the support of the Wharton Interactive Media Initiative, and the third author acknowledges the support of the London Business School Centre for Marketing.

Abstract

New Perspectives on Customer "Death" Using a Generalization of the Pareto/NBD Model

Several researchers have proposed models of buyer behavior in noncontractual settings which assume that customers are "alive" for some period of time, 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 after every purchase occasion. While 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 purchasing while "alive" component, we maintain the assumptions of the Pareto/NBD and BG/NBD models. This results in a model that has the appealing characteristics of the Pareto/NBD with none of its computational burden.

This *periodic death opportunity* (PDO) model allows us to take a closer look at how assumptions about customer death influence model fit. When the time period after which each customer makes his or her dropout decision (which we call *periodicity*) is very small, we show analytically that the PDO model converges to the Pareto/NBD. When the periodicity is longer than the calibration period, the dropout process is "shut off" and the PDO model converges to the NBD model. By systematically varying the periodicity 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; we show this theoretically and our empirical analysis demonstrates the superior performance of the PDO model on two datasets. Finally, we extend the basic model to allow for heterogeneity in periodicity across customers and find that such an extension confirms the results of the basic model.

Keywords: Customer-base analysis, Pareto/NBD, BG/NBD, customer attrition.

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, Morrison, and Colombo 1987), which was the first to capture and exploit a customer death process in such a setting. Applications of the model (e.g., Fader, Hardie and Lee (2005a), 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, Hardie and Lee (2005b), which replaced the continuous-time exponential-gamma process with a discrete-time beta-geometric one. The resulting model, called the BG/NBD, was viewed as a "quick and easy" alternative to the Pareto/NBD, since it offers a much more straightforward parameter estimation process with no substantial loss in the model's fit and forecasting capabilities. But beyond these computational benefits and aggregate indicators of overall model performance, 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, one that combines the "best of both worlds." Our new model is a generalization of the Pareto/NBD yet still enjoys some of the computational benefits of the BG/NBD (e.g., no need to evaluate Gaussian hypergeometric functions in model estimation). Furthermore, this model offers new insights about the death process and some surprising connections between the beta-geometric and exponential-gamma sub-components.

The new framework, called the *periodic death opportunity* (PDO) model, assumes that customers act in accordance with the discrete "coin-flipping" story associated with the beta-geometric, but these "coin flips" arise at periodic intervals in calendar time instead of being

linked to actual purchases (as in the BG/NBD model). Formulated in this manner, the PDO model allows us to take a closer look at how assumptions about customer death influence model fit as well as key managerial inferences/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 periodicity, we effectively vary the customer dropout dynamics. When the periodicity is very large, the dropout component is "shut off" and our model becomes the standard "no death" NBD model. At the other extreme, when the periodicity tends to zero, our discrete-time dropout process converges into the continuous-time dropout process of the Pareto/NBD; we prove this convergence analytically. 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 periodicity, both in the calibration samples and longitudinal holdout periods.

In the next section we formally develop the PDO model. In Section 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 datasets — one using transactions from an online retailer of music CDs, and the other using transactions from a grocery store. In Section 4, we extend the basic model to allow for heterogeneity in the periodicity parameter and estimate this new model on both datasets. Finally, we conclude with a recap of the model, brief discussion of its limitations, and some suggestions for related future research opportunities.

2 Model Development

The periodic death opportunity (PDO) model is based on the following six assumptions:

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

2. While alive, the number of transactions made by a customer follows a Poisson process with transaction rate λ. (This is equivalent to assuming that the interpurchase times are iid exponential with rate λ.)

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

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

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

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

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

5. Heterogeneity in θ follows a beta distribution with pdf

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

6. The transaction rate λ and the dropout probability θ 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

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

while the probability that he 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 β (in place of Assumption 5). This implies that, for a randomly chosen customer,

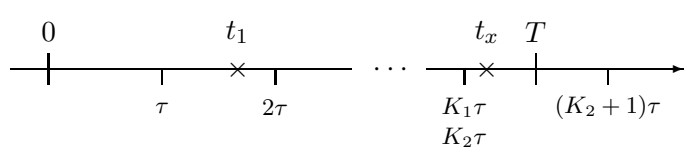
$$E(\Omega \,|\, s, \beta) = \left(\frac{\beta}{s-1}\right) \,, \text{ and} \tag{5}$$

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

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, while in the Pareto/NBD model a customer can die at any point in time. However, one can see that as τ (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 τ 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 $t_1, t_2, \ldots, t_x$; by definition, $t_x = 0$ when $x = 0$. We define $K_1 = \lfloor t_x/\tau \rfloor$ and $K_2 = \lfloor T/\tau \rfloor$.

When $K_1 = K_2$, we have

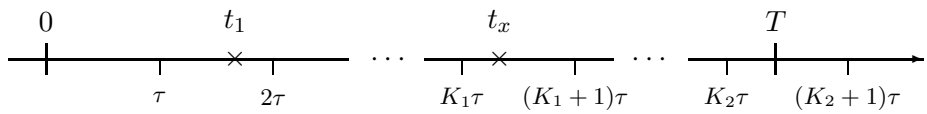


The fact that a purchase occurred at t_x implies the customer must have been "alive" at $K_1\tau$, which occurs with probability $(1-\theta)^{K_1}$. Since $K_1 = K_2$, the customer must still be alive at T.

Given the model assumptions, the likelihood function for this case is

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

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 $(K_1+3)\tau$, ...

- 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 $t_1, \ldots, t_x, T$ 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.)

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

$$L(\lambda, \theta, \tau \,|\, x, t_x, T) = \lambda^x e^{-\lambda T}(1-\theta)^{\lfloor T/\tau \rfloor} + \delta_{\lfloor T/\tau \rfloor > \lfloor t_x/\tau \rfloor} \sum_{j=1}^{\lfloor T/\tau \rfloor - \lfloor t_x/\tau \rfloor} \lambda^x e^{-\lambda(\lfloor t_x/\tau \rfloor + j)\tau} \theta (1-\theta)^{\lfloor t_x/\tau \rfloor + j - 1}. \quad (7)$$

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

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

The five model parameters (r, α, a, b, τ) 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

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

We find the maximum of this function by performing a line search on τ, using standard numerical methods to find the values of r, α, a, b that maximize $LL(r, \alpha, a, b, \tau)$ for a given value of τ.

In Appendix A.1, we show that that as $\tau \to 0$, (8) becomes the likelihood function associated with the Pareto/NBD model. As $\tau \to \infty$, $\lfloor T/\tau \rfloor = 0$ and $\delta_{\lfloor T/\tau \rfloor > \lfloor t_x/\tau \rfloor} = 0$, 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 $\tau > T$.)

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

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

$$E(X(t)\,|\,r,\alpha,a,b,\tau) = \frac{rt}{\alpha}\frac{B(a,b+\lfloor t/\tau\rfloor)}{B(a,b)} + \frac{r\tau}{\alpha}\sum_{j=1}^{\lfloor t/\tau\rfloor} j\frac{B(a+1,b+j-1)}{B(a,b)}\,. \tag{10}$$

  (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

$$\begin{aligned}&P(\Omega > T\,|\,r,\alpha,a,b,\tau,x,t_x,T)\\ &\qquad = \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)}\left(\frac{1}{\alpha+T}\right)^{r+x}\frac{B(a,b+\lfloor T/\tau\rfloor)}{B(a,b)}\Big/L(r,\alpha,a,b,\tau\,|\,x,t_x,T)\,.\end{aligned} \tag{11}$$

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

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

(See Appendix A.2 for the associated derivations.)

3 Empirical Analysis

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

3.1 The CDNOW Dataset

This dataset 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. 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 τ from 0.01 weeks to 40 weeks, 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 1. When τ is very small ($\tau = 0.01$ weeks, $LL = -9{,}594.6$), the fit of the PDO model is almost exactly the same as that of the Pareto/NBD. When τ is large (> 39 weeks in this empirical setting), the PDO model yields the same log-likelihood as that of the NBD model.

What is interesting to note is that the fit of the PDO model can dominate that of the Pareto/NBD. As τ increases from 0.01, the log-likelihood increases from the Pareto/NBD limit to a maximum value of −9,585.3 at $\tau = 3.00$, then starts declining towards the much lower value associated with the NBD. As soon as $\tau > 6$ weeks, the fit of the PDO model is worse than that of the Pareto/NBD. The PDO model (with $\tau = 3$) provides a significant improvement in calibration-period model fit over the Pareto/NBD ($p < .001$ using the likelihood ratio test). The parameter estimates of the best-fitting PDO model and the Pareto/NBD model are reported in Table 1.

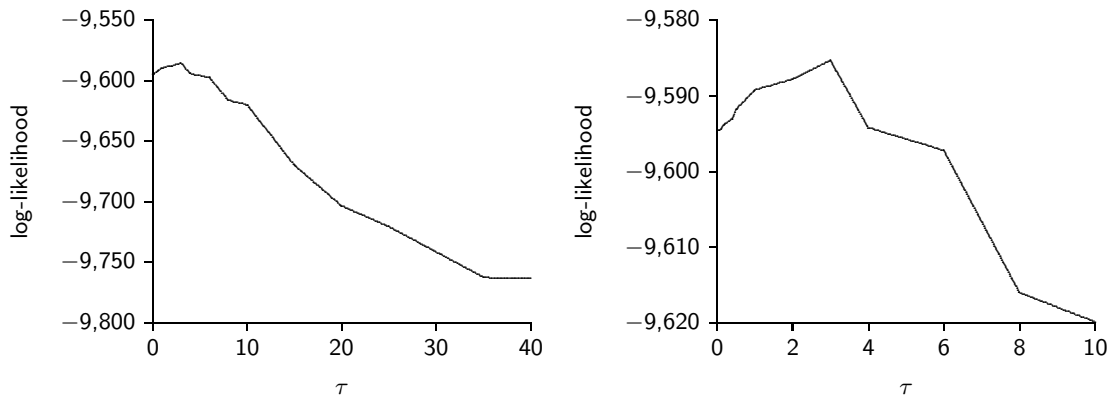


Figure 1: Plot of the PDO model log-likelihood for the CDNOW dataset as a function of the periodicity parameter τ. (The right-hand plot presents a "zoomed in" view for small values of τ.)

	PDO ($\tau = 3$)		Pareto/NBD	
	Estimate	Std. Error	Estimate	Std. Error
r	0.51	0.04	0.55	0.05
α	10.40	0.82	10.58	0.84
a	0.43	0.09		
b	2.61	0.85		
s			0.61	0.19
β			11.67	6.21
LL	−9,585.3		−9,595.0	

Table 1: Parameter estimates and standard errors for the PDO ($\tau = 3$) and Pareto/NBD models for the CDNOW dataset.

To illustrate more clearly the relationship between the Pareto/NBD, the NBD and the PDO models with various values of τ, we present the parameter estimates and other summary statistics for these models in Table 2. Note how the parameters of the PDO model tend towards those of the Pareto/NBD model as $\tau \to 0$ ($a \to s$, $b\tau \to \beta$). This is in accordance with the pattern predicted theoretically in Appendix A.1. 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. 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 to the NBD. Pursuing this logic further, it follows that increases in τ would be associated with declining average transaction rates and increasing median lifetimes. But what is not known is how quickly these statistics will change with τ, 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 τ, but they also reveal some differences. As τ rises from 0 (i.e., Pareto/NBD model) to 3, 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 death process is more sensitive to changes in τ than the mean transaction rates.

	Transaction		Death				
τ	r	α	$a^\dagger$	$b^\dagger$	LL	$E(\Lambda)$	Median(Ω)
(0)	0.55	10.58	0.61	11.67	−9595.0	0.05	25.0
0.01	0.55	10.58	0.61	1167.41	−9,594.6	0.05	25.0
0.10	0.55	10.57	0.60	115.29	−9,594.4	0.05	25.0
1.00	0.54	10.50	0.56	10.90	−9,589.2	0.05	26.5
2.00	0.51	10.16	0.53	5.20	−9,587.8	0.05	26.8
3.00	0.51	10.40	0.43	2.61	−9,585.3	0.05	27.8
4.00	0.50	10.45	0.42	2.02	−9,594.2	0.05	29.8
5.00	0.48	11.26	0.46	2.03	−9,599.0	0.05	30.8
10.00	0.45	10.49	0.26	0.62	−9,619.9	0.04	45.0
20.00	0.40	11.18	0.63	1.51	−9,704.1	0.04	53.8
(∞)	0.39	12.07	--	--	−9,763.7	0.03	--

† s and β in the case of the Pareto/NBD model ($\tau = 0$).

Table 2: Estimation results for the CDNOW dataset for the Pareto/NBD model ($\tau = 0$), the NBD model ($\tau = \infty$) and the PDO model (for various values of τ).

The differences between the Pareto/NBD and PDO ($\tau = 3$) models in terms of the underlying death process are illustrated in Figure 2, which plots $P(\Omega \leq t)$ for both models (as computed using (4) and (6)). The higher median lifetime associated with the PDO ($\tau = 3$) model reflects the slower overall rate of attrition amongst the customer base under this model in this empirical setting.

Coupled with the fact that there is little difference in the mean underlying transaction rates (while alive), $E(\Lambda)$, between the two models, it follows that the PDO ($\tau = 3$) model generates a slightly higher estimate of total repeat sales over time. But is this good or bad? And does the best-fitting PDO specification perform better on holdout data than other specifications ($\tau \neq 3$)? 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 percent error (MAPE) numbers for both cumulative total repeat sales and weekly total repeat sales over weeks 40–78.

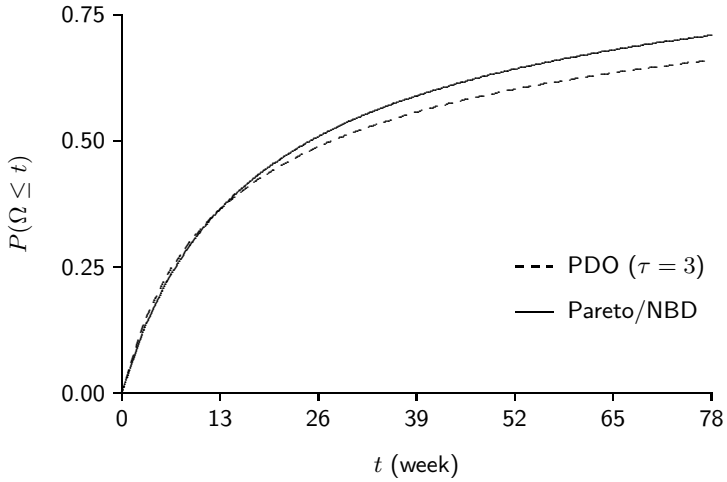


Figure 2: Plot of the probability that a randomly chosen customer will have died by time t ($P(\Omega \leq t)$) under the PDO model with $\tau = 3$ and the Pareto/NBD model.

	MAPE (Weeks 40–78)	
τ	Cumulative	Weekly
(0)	1.35%	20.89%
0.01	1.35%	20.89%
0.10	1.34%	20.53%
1.00	1.14%	20.13%
2.00	1.09%	19.81%
3.00	0.85%	19.06%
4.00	0.85%	19.06%
5.00	0.88%	19.33%
10.00	0.70%	19.18%
20.00	0.92%	19.25%
(∞)	10.37%	36.22%

Table 3: Measures of model forecasting performance for the Pareto/NBD model ($\tau = 0$), the NBD model ($\tau = \infty$) and the PDO model (for various values of τ).

Looking at the cumulative MAPE numbers, we see that all of the PDO models with a finite periodicity 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). But overall, there is strong empirical support for the performance of the PDO model (particularly with $\tau = 3$) 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. But 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 periodic, 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 Dataset

Batislam et al. (2007) used a dataset covering 5,479 individuals at a Turkish grocery store who made their first-ever purchase 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 $-67{,}926.0$ and fitting the NBD model yields $-71{,}000.5$. Fitting the PDO models by varying τ from 0.01 weeks to 80 weeks, we confirm the pattern we observed for the CDNOW dataset — when τ is very small ($\tau = 0.01$ weeks, $LL = -67{,}926.0$), the fit of the PDO model is the same as that of the Pareto/NBD; when τ is large (> 78 weeks in this empirical setting), the PDO model yields the same log-likelihood as that of the NBD model; and the fit of the PDO model dominates that of the Pareto/NBD model for an intermediate range of values of τ (in this case, $\tau \leq 8$ weeks). The values of the log-likelihood function as τ varies from 0.01 weeks to 10 weeks are plotted in Figure 3. The maximum log-likelihood value is achieved when $\tau = 1.00$ and is $-67{,}749.8$. Compared to the Pareto/NBD model, this is a highly significant improvement in model fit — an improvement of 176 log-likelihood points at the cost of just one additional parameter. In Table 4, we present the parameter estimates of the best-fitting PDO model and the Pareto/NBD model.

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 τ varies, and the same qualitative relationships between these models and the Pareto/NBD

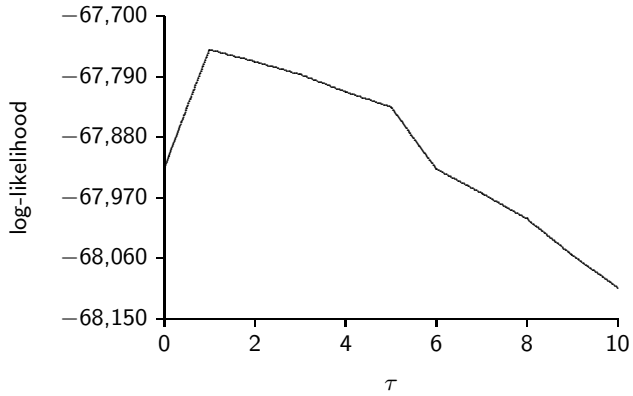


Figure 3: Plot of the PDO model log-likelihood for the Grocery dataset as a function of the periodicity parameter τ.

	PDO ($\tau = 1$)		Pareto/NBD	
	Estimate	Std. Error	Estimate	Std. Error
r	0.46	0.01	0.48	0.02
α	4.38	0.18	4.38	0.16
a	0.62	0.03		
b	22.19	0.02		
s			0.57	0.05
β			17.64	2.87
LL	−67,749.8		−67,925.8	

Table 4: Parameter estimates and standard errors for the PDO ($\tau = 1$) and Pareto/NBD models for the Grocery dataset.

model. For instance, for the best-fitting PDO model (with $\tau = 1$), 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 ($\tau = 1$) 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 dataset, and we observe similar relationships between the Pareto/NBD, NBD and various PDO models in this dataset as we did in the CDNOW dataset.

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 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 dataset; its log-likelihood value of −9,582.4, as reported in Fader et al. (2005b), is slightly better than that of the PDO model with $\tau = 3$ (and therefore substantially better than that of the regular Pareto/NBD). However, for the Grocery dataset, the BG/NBD model ($LL = -68{,}007.0$) significantly underperforms both the PDO model and the Pareto/NBD model.[1] 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 dataset, the BG/NBD model's cumulative MAPE is 2.6% and weekly MAPE is 19.4% (both slightly worse). For the Grocery dataset, the BG/NBD model's cumulative MAPE is 8.8% (much worse) and weekly MAPE is 4.0% (slightly worse).

But beyond the empirical evidence, it is important to emphasize the theoretical benefits of the PDO relative to the BG/NBD. The BG/NBD has no direct connection to the Pareto/NBD (although it does nest the basic NBD), and there is no way to equate the parameters across these different specifications. These conceptual benefits may give the edge to the PDO framework, particularly for researchers with a focal interest in the nature of the death process.

4 Heterogeneity in τ

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

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

[1] The BG/NBD model parameter estimates are $\hat{r} = 0.24, \hat{\alpha} = 4.41, \hat{a} = 0.79, \hat{b} = 2.43$ for the CDNOW dataset, and $\hat{r} = 0.28, \hat{\alpha} = 2.34, \hat{a} = 0.40, \hat{b} = 2.09$ for the Grocery dataset.

(Using other heterogeneity distributions, such as a lognormal 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 τ, 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 MCMC methods for estimating this heterogeneous PDO (henceforth HPDO) model; see Appendix A.3 for details.

We estimate the HPDO model on both datasets. Our inferences are based on 50,000 iterations, following a burn-in of 20,000 iterations. We use different starting values and confirm convergence to the same values every time for both datasets. The resulting parameter estimates are presented in Table 5. (The associated log-marginal density numbers are −8,577.1 and −63,746.2 for the CDNOW and Grocery datasets, respectively.)

	CDNOW		Grocery	
	Posterior mean	95% Interval	Posterior mean	95% Interval
r	0.41	[0.408, 0.413]	0.35	[0.349, 0.351]
α	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]
ρ	0.12	[0.118, 0.123]	0.21	[0.207, 0.214]

Table 5: Parameter estimates for the HPDO model for the CDNOW and Grocery datasets.

Figure 4 shows the distribution of τ for both datasets. Clearly, there is significant heterogeneity in τ in both datasets. Note that the expected value of τ is larger for the CDNOW dataset as compared to the Grocery dataset; this is in line with the fact that the value of τ for the best-fitting PDO model (i.e., without heterogeneity in τ) is larger for the CDNOW dataset as compared to the Grocery dataset. However, while 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 τ. For the CDNOW

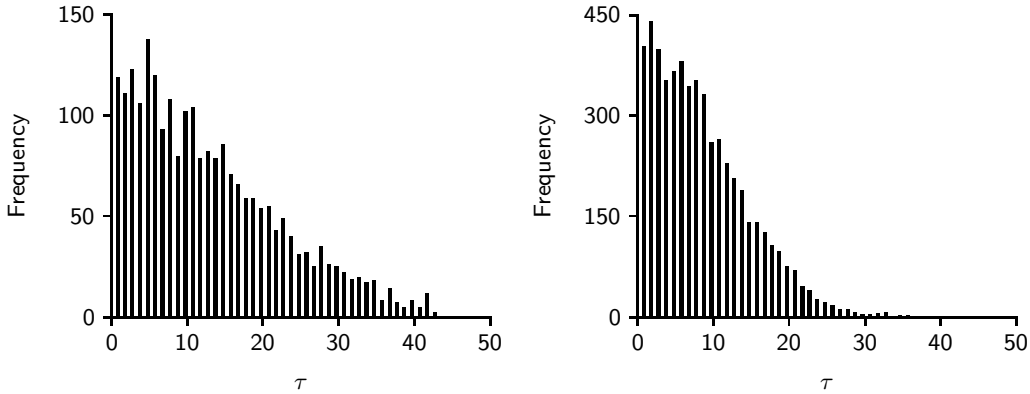


Figure 4: Histogram of posterior draws of τ across customers from one iteration after convergence for the CDNOW dataset (left) and the Grocery dataset (right).

dataset, for weeks 40–78, the cumulative MAPE is 1.92% and the weekly MAPE is 20.55%. For the Grocery dataset, for weeks 78–91, the cumulative MAPE is 1.5% and the weekly MAPE is 8.1%. In this respect, 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 a great deal of value to future researchers, particularly if they want to test theories about how the nature of the dropuout process varies across customers.

5 Discussion and Conclusions

We have proposed the periodic death opportunity (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.

But while 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 view the Pareto/NBD as a natural benchmark which is still very useful when the manager's primary goal is forecasting purchases, as opposed to a principal focus on the death process, per se. Although in our datasets 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.

Besides improvements in model performance and parameter inferences, the PDO model might offer other managerial benefits as well. For instance, while we hesitate to suggest that the periodic death opportunity "story" is a true representation of customer decision-making, the τ parameter may serve as an indicator of any "re-evaluation" decision that each customer may be undertaking. In this sense, it might offer useful guidance about customer contact strategies — for instance, it might help managers determine which customers should be "reminded" more often, and which ones can be left alone for longer periods of time. It is easy to envision such allocation decisions as an explicit component as part of a broader program to prevent customer attrition.

Finally, a logical extension of our hierarchical Bayesian estimation procedure would be the use of covariates to help explain how the death process varies across customers — something that is already being investigated for the regular Pareto/NBD model by several researchers (e.g., Abe 2009, Fader and Hardie 2007) — and perhaps over time. This could help managers manage customer attrition in an effective manner. But these issues 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 and its computational simplicity.

Appendix

A.1 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 $\tau \to 0$.

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 λ and μ is distributed gamma with parameters (r, α) and (s, β), respectively, it follows that the likelihood function for a randomly-chosen customer is found by solving

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

Our proof is based on showing that

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

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

This proof will make use of the following results:

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

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

$$\lim_{\tau \to 0} \frac{\Gamma(s + \beta/\tau)}{\Gamma(\beta/\tau)} \tau^s = \beta^s \,. \tag{A4}$$

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

Noting that

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

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

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

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

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

where

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

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

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

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

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

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

and

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

It follows that

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

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

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

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 $\tau \to 0$ and the Pareto/NBD. We use this fact somewhat liberally

to "prove" that the two models are equivalent under this special condition ($\tau \to 0$). (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 $\tau \to 0$.)

A.2 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 λ, it follows from the assumption of Poisson purchasing that $E[X(t)]$ is simply λt if the customer is active at t, $\lambda\tau$ if the customer dies at τ, $2\lambda\tau$ if the customer dies at 2τ, $3\lambda\tau$ if the customer dies at 3τ, ..., 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 τ, 2τ, ..., 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 λ and θ, (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 he was alive at $K_2\tau$. Referring back to our derivation of the individual-level likelihood function, (7)), the application of Bayes theorem gives us

$$P(\Omega > T \,|\, \lambda, \theta, \tau, x, t_x, T) = \frac{\lambda^x e^{-\lambda T} (1-\theta)^{\lfloor T/\tau \rfloor}}{L(\lambda, \theta, \tau \,|\, x, t_x, T)} \,. \tag{A10}$$

(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 λ and θ is given by

$$f(\lambda, \theta \,|\, r, \alpha, a, b, \tau, , x, t_x, T) = \frac{L(\lambda, \theta, \tau \,|\, x, t_x, T) f(\lambda \,|\, r, \alpha) f(\theta \,|\, a, b)}{L(r, \alpha, a, b, \tau \,|\, x, t_x, T)} \,. \tag{A11}$$

Taking the expectation of (A10) 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 $E(X(T, T+t) \mid x, t_x, T)$, 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 λ, it follows from the assumption of Poisson purchasing that the expected number of purchases in $(T, T+t]$ is simply λt 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 $(\lfloor T/\tau \rfloor + 2)\tau$, ..., and $\lambda(\lfloor (T+t)/\tau \rfloor \tau - T)$ if the customer dies at $\lfloor (T+t)/\tau \rfloor \tau$. Multiplying these quantities by the probability that the customer dies at $(\lfloor T/\tau \rfloor + 1)$, $(\lfloor T/\tau \rfloor + 2)$, ..., gives us

$$E(X(T, T+t) \mid \lambda, \theta, \tau, \Omega > T) = \lambda t (1-\theta)^{\lfloor (T+t)/\tau \rfloor - \lfloor T/\tau \rfloor} + \sum_{j=1}^{\lfloor (T+t)/\tau \rfloor - \lfloor T/\tau \rfloor} \lambda[(\lfloor T/\tau \rfloor + j)\tau - T]\theta(1-\theta)^{j-1} . \quad \text{(A12)}$$

Taking the expectation of the product of (A10) and (A12) over the joint posterior distribution of λ and θ, (A11), gives us (12).

A.3 MCMC Procedure for HPDO Model

Customer i with periodicity τ_i makes his dropout decision at $\tau_i, 2\tau_i, 3\tau_i, \ldots$. Whether the customer dropped out or not and, if so, when he 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 $\psi_i \in \{\lfloor t_{x_i}/\tau_i \rfloor + 1, \ldots, \lfloor T_i/\tau_i \rfloor\}$, $Z_i = 0$.

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,

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

The parameters of the heterogeneity distributions for λ, θ and τ specified earlier act as priors for λ_i, θ_i and τ_i. 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 λ_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,

$$\lambda_i \,|\, r, \alpha, \tau_i, x_i, T_i, z_i, \psi_i \sim \text{gamma}(r + x_i - 1, \alpha + z_i T_i + (1-z_i)\psi_i \tau_i) . \tag{A13}$$

- The conditional posterior distribution of θ_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 $\theta_i^{a - z_i}(1-\theta_i)^{b - 1 + z_i \lfloor T_i/\tau_i \rfloor + (1-z_i)(\psi_i - 1)}$. Therefore,

$$\theta_i \,|\, a, b, \tau_i, T_i, z_i, \psi_i \sim \text{beta}(a + 1 - z_i, b + z_i \lfloor T_i/\tau_i \rfloor + (1-z_i)(\psi_i - 1)) . \tag{A14}$$

- The conditional posterior distribution of τ_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

$$f(\tau_i \,|\, m, \rho, \tau_i, T_i, z_i, \psi_i) \propto \tau_i^{m-1} e^{-[\rho \tau_i + \lambda_i (1-z_i)\psi_i \tau_i]} (1-\theta_i)^{z_i \lfloor T_i/\tau_i \rfloor} . \tag{A15}$$

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

- Recalling (A10),

$$P(Z_i = 1 \,|\, \lambda_i, \theta_i, \tau_i t_{x_i}, T_i) = \frac{\mathcal{B}_1}{\mathcal{B}_1 + \delta_{\lfloor T_i/\tau_i \rfloor > \lfloor t_{x_i}/\tau_i \rfloor} \mathcal{B}_2} \tag{A16}$$

  where

$$\mathcal{B}_1 = e^{-\lambda_i T_i}(1-\theta_i)^{\lfloor T_i/\tau_i \rfloor} \text{ and } \mathcal{B}_2 = \sum_{j=1}^{\lfloor T_i/\tau_i \rfloor - \lfloor t_{x_i}/\tau_i \rfloor} e^{-\lambda_i(\lfloor t_{x_i}/\tau_i \rfloor + j)\tau_i} \theta_i (1-\theta_i)^{\lfloor t_{x_i}/\tau_i \rfloor + j - 1} .$$

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

- If $Z_i = 0$, then we also generate an integer $\psi_i \in \{\lfloor t_{x_i}/\tau_i \rfloor + 1, \ldots, \lfloor T_i/\tau_i \rfloor\}$, 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 ψ_i,

$$P(\Psi_i = \psi_i \mid \lambda_i, \theta_i, \tau_i, t_{x_i}, T_i) \propto e^{-\lambda_i \psi_i \tau_i} \theta_i (1-\theta_i)^{\psi_i - 1} \,. \tag{A17}$$

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

$$f(r, \alpha \mid \boldsymbol{\lambda}) \propto \left\{ \prod_{i=1}^{I} f(\lambda_i \mid r, \alpha) \right\} f(r) f(\alpha) \,, \tag{A18}$$

$$f(a, b \mid \boldsymbol{\theta}) \propto \left\{ \prod_{i=1}^{I} f(\theta_i | a, b) \right\} f(a) f(b) \,, \tag{A19}$$

$$f(m, \rho \mid \boldsymbol{\tau}) \propto \left\{ \prod_{i=1}^{I} f(\tau_i | m, \rho) \right\} f(m) f(\rho) \,. \tag{A20}$$

where $\boldsymbol{\lambda}$, $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$ denote the current vectors of the individual-level parameters. We use weakly informative gamma hyperpriors for r, α, a, b, m and ρ. (The choice of these priors does not influence our final estimates (except for ensuring that they are positive) because the two datasets we use have a large number of individuals in them.) Metropolis-Hastings methods are used to sample the pairs (r, α), (a, b) and (m, ρ).

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

1. Set initial values for r, α, a, b, m, ρ and $\lambda_i, \theta_i, \tau_i, z_i, \psi_i, \forall\, i = 1, \ldots, I$.

2. Iterate until convergence:

   (a) For each customer i, sample λ_i, θ_i and τ_i using (A13), (A14) and (A15), respectively.

   (b) For each customer i, generate z_i using (A16). If $z_i = 0$, generate ψ_i using (A17). (This is the data augmentation step.)

   (c) Sample the pairs (r, α), (a, b) and (m, ρ) using (A18), (A19), and (A20), respectively.

References

Abe, Makoto (2009), ""Counting Your Customers" One by One: A Hierarchical Bayes Extension to the Pareto/NBD Model," *Marketing Science*, forthcoming.

Abramowitz, Milton and Irene A. Stegun (eds.) (1972), *Handbook of Mathematical Functions*, New York: Dover Publications.

Batislam, Emine P., Meltem Denizel and Alpay Filiztekin (2007), "Empirical Validation and Comparison of Models for Customer Base Analysis," *International Journal of Research in Marketing*, **24** (September), 201–209.

Fader, Peter S. and Bruce G. S. Hardie (2005), "A Note on Deriving the Pareto/NBD Model and Related Expressions," `<http://brucehardie.com/notes/009/>`.

Fader, Peter S. and Bruce G. S. Hardie (2007), "Incorporating Time-Invariant Covariates into the Pareto/NBD and BG/NBD Models." <`http://brucehardie.com/notes/019/`>

Fader, Peter S., Bruce G. S. Hardie and Ka Lok Lee (2005a), "RFM and CLV: Using Iso-Value Curves for Customer Base Analysis," *Journal of Marketing Research*, **42** (November), 415–430.

Fader, Peter S., Bruce G. S. Hardie and Ka Lok Lee (2005b), ""Counting Your Customers" the Easy Way: An Alternative to the Pareto/NBD Model," *Marketing Science*, **24** (Spring), 275–284.

Gupta, Sunil and Donald G. Morrison (1991), "Estimating Heterogeneity in Consumers' Purchase Rates," *Marketing Science*, **10** (Summer), 264–269.

Reinartz, Werner and V. Kumar (2000), "On the Profitability of Long-Life Customers in a Noncontractual Setting: An Empirical Investigation and Implications for Marketing," *Journal of Marketing*, **64** (October), 17–35.

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

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

Tanner, Martin A. and Wing Hung Wong (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, **82** (June), 528–540.