ESTIMATING CLV USING AGGREGATED DATA: THE *TUSCAN LIFESTYLES* CASE REVISITED

PETER S. FADER, BRUCE G. S. HARDIE, AND KINSHUK JERATH

The *Tuscan Lifestyles case* (Mason, 2003) offers a simple twist on the standard view of how to value a newly acquired customer, highlighting how standard retention-based approaches to the calculation of expected customer lifetime value (CLV) are not applicable in a noncontractual setting. Using the data presented in the case (a series of annual histograms showing the aggregate distribution of purchases for two different cohorts of customers newly "acquired" by a catalog marketer), it is a simple exercise to compute an estimate of "expected 5 year CLV." If we wish to arrive at an estimate of CLV that includes the customer's "life" beyond five years or are interested in, say, sorting out the purchasing process (while "alive") from the attrition process, we need to use a formal model of buying behavior that can be applied on such coarse data.

To tackle this problem, we utilize the Pareto/NBD model developed by Schmittlein, Morrison, and Colombo (1987). However, existing analytical results do not allow us to estimate the model parameters using the data summaries presented in the case. We therefore derive an expression that enables us to do this. The resulting parameter estimates and subsequent calculations offer useful insights that could not have been obtained without the formal model. For instance, we were able to decompose the lifetime value into four factors, namely purchasing while active, dropout, surge in sales in the first year and monetary value of the average purchase. We observed a kind of "triple jeopardy" in that the more valuable cohort proved to be better on the three most critical factors.

PETER S. FADER

is Frances and Pei-Yuan Chia

Professor of Marketing at

The Wharton School of the

University of Pennsylvania;

e-mail: faderp@wharton. upenn.edu

• •••

BRUCE G. S. HARDIE

is Associate Professor of Marketing at

the London Business School;

e-mail: bhardie@london.edu

• •••

KINSHUK JERATH

is a doctoral candidate in the

Marketing Department

at The Wharton School of the

University of Pennsylvania;

e-mail: kinshuk@wharton. upenn.edu

• •••

INTRODUCTION

Most standard introductions to the notion of customer lifetime value (CLV) center around a formula similar to

$$CLV = \sum_{t=0}^{\infty} m \frac{r^t}{(1 + d)^t} \qquad (1)$$

(where m is the net cash flow per period, r is the retention rate, and d is the discount rate) and claim that this is the appropriate starting point for the calculation of lifetime value.

However, such an expression is not applicable in many business settings, particularly those that can be viewed as *noncontractual*. A defining characteristic of a noncontractual setting is that the time at which a customer becomes inactive is unobserved by the firm; customers do not notify the firm "when they stop being a customer. Instead they just silently attrite" (Mason, 2003, p. 55). This is in contrast to a contractual setting, where the time at which the customer becomes inactive *is* observed (e.g., when the customer fails to renew his or her subscription, or contacts the firm to cancel his or her contract). When the point at which the customer disappears is not observed, we cannot meaningfully utilize notions such as "retention rates" and therefore formulae along the lines of (1) are not appropriate. We can, however, capture the "silent attrition" phenomenon by using a probabilistic dropout process for each customer. We can define the "survival probability," $S(t)$, for each customer at a given time t, (i.e., the probability that the customer is "alive" at t). This leads to the following definitional expression for expected CLV

$$E(CLV) = \sum_{t=0}^{\infty} \frac{E[v(t)]S(t)}{(1 + d)^t}, \qquad (2)$$

where $E[v(t)]$ is the expected value (or net cash flow) of the customer at time t (if active). The challenge is to operationalize (2) in any given setting. (See, for example, Fader, Hardie, & Lee, 2005.)

One example of a noncontractual business setting is presented in the *Tuscan Lifestyles* case (Mason, 2003). This case provides a summary of repeat buying behavior for a group of 7,953 new customers over a 5-year period beginning immediately after their first-ever purchase. These data are presented in Table 1.

We have five annual histograms for two groups of customers—the first comprising 4,657 customers with an initial purchase below $50 and the second comprising 3,296 customers with an initial purchase greater than or equal to $50. (Note that the "years" do not refer to calendar time, but reflect time since initial purchase.) The task in the case is to compute the value of a new *Tuscan Lifestyles* customer (i.e., estimate CLV).

Besides highlighting the inapplicability of the standard CLV formula, the *Tuscan Lifestyles* case also brings forth another important issue—the practical limitations of obtaining detailed transaction-level data. Even though many researchers have developed general frameworks and specific methods for modeling customer lifetime value (CLV) using disaggregate data, few have carefully considered the difficult realities of firms' abilities (or inabilites) to deal with customer-level data. While many reporting systems are able to create simple data summaries for a fixed period of time (e.g., an annual histogram of number of purchases), the process of extracting raw individual-level data can be a time-consuming task (especially if the information technology group is not directly involved with the project).

High-profile stories on data loss, such as the 2005 loss of tapes containing information on 3.9 million Citigroup customers or the August 2006 loss of a computer containing data on 26.5 million veterans by a Department of Veterans Affairs (VA) subcontractor, have justifiably made a number of companies wary of releasing customer-level data. Coupled with rising consumer concerns about privacy, this has motivated a major research stream in information systems called *privacy preserving data mining* (e.g., Agrawal & Srikant, 2000). However, the process of "anonymizing" customer data can be challenging, making it only more difficult for marketing to get the information technology group to extract the data required for modeling. Furthermore, there are growing concerns regarding the extent to which privacy is actually preserved in anonymized databases (Mielikäinen, 2004).

Moreover, data protection laws in many countries (particularly in Europe) complicate the process of transferring raw data to the analyst (Carey, 2004; Singleton, 2004), even to the extent of there being

TABLE 1	*Tuscan Lifestyles* Data: Number of Purchases per Year (Not Including Initial Purchase) for Each of Two Cohorts Grouped by Size of Initial Purchase

	<$50 COHORT						≥$50 COHORT				
	YEAR						YEAR				
# ORDERS	1	2	3	4	5	# ORDERS	1	2	3	4	5
0	611	2739	3687	3730	3837	0	421	1643	2430	2535	2673
1	3508	1441	671	661	626	1	2354	1120	562	548	463
2	416	332	207	185	141	2	397	354	214	151	120
3	94	100	68	56	38	3	91	121	53	37	30
4	21	30	19	14	10	4	20	53	27	20	7
5	7	9	1	9	3	5	6	12	6	3	2
6		3	2	2		6	5	5	2	2	1
7		1	1			7	1	3	1		
8		1			2	8	1				
9		1				9			1		
10			1			10					
11						11					
12						12					
13						13		1			

partial bans on transborder data flows. Given general outsourcing trends, these laws can create real barriers to the implementation of models built on individual-level data.

For these reasons, it is often much harder to implement CLV models than a casual reading of the marketing literature might suggest. When the transaction data are summarized across customers, such as in Table 1, the concerns raised above evaporate. But this leads us back to the issues at the heart of this paper: how to compute CLV from such data.

Using the data provided above, we can easily arrive at an estimate of "expected 5 years CLV." But what about the customer's "life" beyond 5 years? And what if we wish to know more than just the mean purchase rate? For instance, suppose we are interested in sorting out the purchasing process (while "alive") from the attrition process? Any serious examination of CLV—and any corporate program that relies on it—should consider such questions. Unfortunately they cannot be answered using these data alone. This situation is not unique; other researchers (e.g., Berger, Weinberg, &

Hanna, 2003) have also relied on aggregate data and therefore are subject to similar limitations.

Thus, instead of using relatively simple "accounting" methods to tally up the past value of each customer segment, we need a formal model to capture the underlying purchase patterns and then project them out to future periods. This is where a stochastic model of customer behavior comes in. Such a model posits latent probabilistic processes which are presumed to underlie the observable behavior of each customer. In the CLV setting, we need to develop a probabilistic model that takes into account three distinct (but possibly inter-related) processes: (1) the purchase frequency of a customer while active, (2) the attrition in the customer base over time, and (3) the monetary value of the purchases. Such a model can be fit using recorded data for the early activity of the customer base and future purchases can then be predicted. While the model is initially conceptualized at the level of the individual customer, it is then aggregated across a population of heterogeneous customers and estimated using data at the segment level or across the entire customer base (while still recognizing the underlying sources of heterogeneity).

In this paper, we invoke the Pareto/NBD framework (Schmittlein, Morrison, & Colombo, 1987), a parsimonious model of repeat buying behavior in a noncontractual setting that provides excellent predictive power using limited summary information about each customer. However, in its original form, the parameters of the Pareto/NBD cannot be be estimated using aggregated data of the form given in Table 1. In the next section, we derive an expression that enables us to estimate the model parameters using such data. We then fit the model to the data and examine the key question: "What is a new *Tuscan Lifestyles* customer worth?"

MODEL DEVELOPMENT

The Pareto/NBD is a powerful stochastic model of purchasing in a noncontractual setting. It starts by assuming that a customer's relationship with the firm has two phases: he or she is "alive" for an unobserved period of time, then becomes permanently inactive. While alive, the customer is assumed to purchase "randomly" around his or her mean transaction rate. As such, a customer's sequence of purchases can appear to be somewhat lumpy/uneven at times, even though the unobserved buying rate is constant. The unobserved time at which the customer becomes permanently inactive is also the outcome of a probabilistic process governed by a dropout rate specific to the customer. We assume that customers are heterogeneous: both the transaction rates and dropout rates vary from person to person.

More formally, the assumptions of the model are as follows:

i. Customers go through two stages in their "lifetime" with a specific firm: They are alive for some period of time, then become permanently inactive.

ii. While alive, the number of transactions made by a customer follows a Poisson process with transaction rate λ.

iii. A customer's unobserved "lifetime" of length ω (after which he is viewed as being permanently inactive) is exponentially distributed with dropout rate μ.

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

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

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

Schmittlein, Morrison, and Columbo (1987) present a careful justification of these assumptions. The empirical validations of the model presented in Schmittlein and Peterson (1994) and Fader, Hardie, and Lee (2005) provide further (indirect) support for these assumptions.

Given these assumptions, it possible to derive expressions for expected purchasing, mean (or median) lifetime, expected CLV, and so on. In order to compute these quantities, we need to know the values of the four model parameters: r, α (which characterize the distribution of transaction rates across the customer base) and s, β (which characterize the distribution of dropout rates across the customer base).

If we start by assuming that we know the exact timing of all the transactions associated with each customer, it turns out that we can estimate the four model parameters using a likelihood function that only requires "recency" (the time of the last purchase) and "frequency" (how many purchases occurred in a given time period) information for each customer. However, in many situations we do not have access to such data; for example, we may only have summaries such as those given in Table 1. The problem with such a data structure is that any longitudinal information about an individual customer is lost. Suppose someone made two repeat purchases in year 1; we do not know how many purchases they made in years 2–5. Does this mean we cannot apply the Pareto/NBD model?

If we reflect on the above model assumptions, we see that they tell a "story" about customer behavior that is not at all related to the nature of the data that might be available to estimate the model parameters. (This is the hallmark of a stochastic model—tell the story first, then deal with data issues later.)

Let the random variable $X(t, t + 1)$ denote the number of transactions observed in the time interval $(t, t + 1]$. Referring back to the $< \$50$ group in Table 1, we see that $X(0, 1) = 0$ for 611 people, $X(1, 2) = 1$ for 1,441 people, and so on. If we can derive an expression for

$P(X(t, t + 1) = x)$ as implied by the Pareto/NBD model assumptions, we can then use it as a means of estimating the four model parameters given the data in Table 1.

Suppose we know an individual's unobserved latent characteristics λ and μ. For $x > 0$, there are two ways x purchases could have occurred in the interval $(t, t + 1]$:

i. the individual was alive at t and remained alive through the whole interval, making x purchases during this interval, or

ii. the individual was alive at t but "died" at some point ω ($< t + 1$), making x purchases in the interval $(t, \omega]$.

For the case of $x = 0$, there is an additional reason as to why no purchases could have occurred in the interval $(t, t + 1]$: The individual was "dead" at t. Given model assumptions (ii) and (iii), we can derive the following expression for the probability of observing x purchases in the interval $(t, t + 1]$, conditional on λ and μ:

$$P(X(t, t + 1) = x \,|\, \lambda, \mu)$$

$$= \delta_{x=0}\left[1 - e^{-\mu t} \right] + \frac{\lambda^x e^{-\lambda} e^{-\mu(t+1)}}{x!}$$

$$+ \left(\frac{\lambda}{\lambda + \mu} \right)^x \left(\frac{\mu}{\lambda + \mu} \right) e^{-\mu t} \qquad (3)$$

$$- \left(\frac{\lambda}{\lambda + \mu} \right)^x \left(\frac{\mu}{\lambda + \mu} \right) e^{-\lambda} e^{-\mu(t+1)}$$

$$\times \sum_{i=0}^{x} \frac{[(\lambda + \mu)]^i}{i!},$$

where $\delta_x = 0$ equals 1 if $x = 0$, 0 otherwise.

In reality, we never know an individual's latent characteristics; we therefore remove the conditioning on λ and μ by taking into account the distributions of the transaction and dropout rates, giving us

$$P(X(t, t + 1) = x \,|\, r, \alpha, s, \beta)$$

$$= \delta_{x=0}\left[1 - \left(\frac{\beta}{\beta + t} \right)^s \right] + \frac{\Gamma(r + x)}{\Gamma(r)x!} \left(\frac{\alpha}{\alpha + 1} \right)^r \left(\frac{1}{\alpha + 1} \right)^x \left(\frac{\beta}{\beta + t + 1} \right)^s$$

$$+ \alpha^r \beta^s \frac{B(r + x, s + 1)}{B(r, s)} \left\{ B_1 - \sum_{i=0}^{x} \frac{\Gamma(r + s + i)}{\Gamma(r + s)} \frac{1}{i!} B_2 \right\} \qquad (4)$$

where

$$B_1 = \begin{cases} {}_2F_1\left(r + s, s + 1; r + s + x + 1; \dfrac{\alpha - (\beta + t)}{\alpha} \right) \Big/ \alpha^{r+s} & \text{if } \alpha \geq \beta + t \\[2ex] {}_2F_1\left(r + s, r + x; r + s + x + 1; \dfrac{\beta + t - \alpha}{\beta + t} \right) \Big/ (\beta + t)^{r+s} & \text{if } \alpha \leq \beta + t \end{cases} \qquad (5)$$

$$B_2 = \begin{cases} {}_2F_1\left(r + s + i, s + 1; r + s + x + 1; \dfrac{\alpha - (\beta + t)}{\alpha + 1} \right) \Big/ (\alpha + 1)^{r+s+i} & \text{if } \alpha \geq \beta + t \\[2ex] {}_2F_1\left(r + s + i, r + x; r + s + x + 1; \dfrac{\beta + t - \alpha}{\beta + t + 1} \right) \Big/ (\beta + t + 1)^{r+s+i} & \text{if } \alpha \leq \beta + t \end{cases} \qquad (6)$$

and $_2F_1(\cdot)$ is the Gaussian hypergeometric function. (A step-by-step derivation of (3)–(6) is presented in the technical appendix.)

Given data summaries of the form presented in Table 1, we can estimate the four Pareto/NBD model parameters via the method of maximum likelihood in the following manner. Suppose we have a sample of T period-specific histograms that give us the distribution of the number of purchases across a fixed set of customers in each period (of equal length). Let $n_{x,t}$ be the number of people who made x purchases in the tth period. (Referring back to the $< \$50$ group in Table 1, $T = 5$, $n_{0,1} = 611$, $n_{1,2} = 1{,}441$, and so on.) The sample log-likelihood function is given by

$$LL(r,\alpha,s,\beta) = \sum_{t=0}^{T-1} \sum_{x=0}^{\infty} n_{x,t} \ln\left[P(X(t, t+1) = x | r, \alpha, s, \beta)\right].$$

(7)

This can be maximized using standard numerical optimization routines.

MODEL ESTIMATION RESULTS

We first applied the basic model to the data in Table 1. Using (7), we obtained the maximum likelihood estimates of the model parameters for each of the two cohorts of customers. We then generated the purchase histograms for the first 5 years exactly as given and compared these generated histograms to the original data. Looking closely at the raw data (Table 1), we can see that there is a large number of customers making one repeat purchase in the first year (for both cohorts). This number then drops sharply in the second year, after which it declines smoothly. On the other hand, the numbers of customers making more than one repeat purchase do not show any sharp variations. While the Pareto/NBD model is very flexible, it is not flexible enough to capture this year 1 aberration—not only did it miss the spike at $x = 1$, but in an attempt to capture this surge in the first year, the predictions for the later years were off as well.

Many researchers would be tempted to propose a more complicated story of buyer behavior in order to accommodate this aberration. However, inspired by Fader and Hardie (2002), we accommodate this year 1 deviation simply by adding a "spike" in the probability of making one repeat purchase in the first year. (In the absence of adequate knowledge of the true, underlying data generation process, one can ex post consider the characteristics of the collected data that might have led to such patterns. For instance, *Tuscan Lifestyles* might have offered a coupon to its new customers that would expire 1 year after their initial purchases.)

More formally, we add a single parameter for each group of customers to address this problem. We assume that, within the first year after making the initial purchase, a "hard-core" fraction π of the customers in the cohort make exactly one repeat purchase that year, with the remaining fraction $1 - \pi$ purchasing according to the basic Pareto/NBD process.

Under this augmented model, the probability that a customer makes x purchases in the $(t + 1)$th period is

$$P(X(t, t+1) = x | r, \alpha, s, \beta, \pi) =$$
$$\begin{cases} \pi + (1 - \pi)P_{PNBD}\left(X(t, t+1) = x\right) & \text{if } t = 0 \text{ and } x = 1 \\ (1 - \pi)P_{PNBD}(X(t, t+1) = x) & \text{if } t = 0 \text{ and } x \neq 1 \\ P_{PNBD}(X(t, t+1) = x) & \text{if } t > 0 \end{cases}$$

(8)

where $P_{PNBD}(\cdot)$ is the basic Pareto/NBD Probability given in (4). From here on, we will only use this "Pareto/NBD with spike" model for the results and analysis that follow. Applying the maximum likelihood estimation procedure in the same manner as for the basic Pareto/NBD model, we obtain the parameter estimates for each group as reported in Table 2.

These parameters can be interpreted by plotting the various mixing distributions that they characterize. Figure 1 shows how the transaction rates (λ) and the dropout rates (μ) vary across the members of each cohort. The high values of r and α indicate that there

| TABLE 2	Parameter Estimates by Cohort				
COHORT	τ	α	S	β	π
$<\$50$	32.83	37.21	12.13	37.74	0.63
$\geq\$50$	148.11	142.07	29.00	92.26	0.57

Transaction Rates
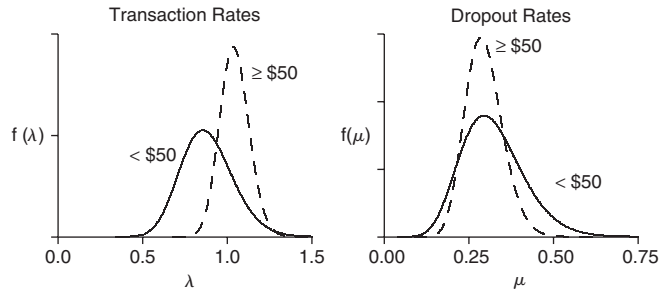
Dropout Rates

FIGURE 1

Heterogeneity in Transaction and Dropout Rates for the Two Cohorts

is relatively little heterogeneity in the underlying transaction rate λ. Similarly, the high values of s and β indicate that there is little heterogeneity in the underlying dropout rate μ. Nevertheless, there are some noteworthy differences across the two groups. It is clear that the transaction rates tend to be higher for the $\geq \$50$ group, albeit with a lower variance. The dropout rates are much closer across the two groups, but they tend to be slightly higher for the $< \$50$ group. Finally, the π parameters indicate that a hard-core of roughly 60% of the customers make just one repeat purchase in the first year, and this proportion is about the same for each group.

These "stories" about the underlying behavioral tendencies within each segment seem to be plausible (and managerially interesting). Looking at the raw data alone provides no intuition about the interplay between the flow of transactions for active customers and the difference in dropout tendencies both within and across each of the customer groups.

Even stronger support for the model is offered in Figure 2; for each of the 5 years, we compare the actual distribution of the number of transactions per year with the corresponding fitted distribution (computed using (8) and the parameter estimates given in Table 2), by cohort. Table 3 shows the mean absolute percentage error (MAPE) between the actual and predicted numbers for the 5 years, individually and across all the years (combined). It is quite remarkable to see how well a five-parameter model can capture the different shapes that are seen within each set of histograms. More importantly, the model seems to do an excellent job of following the systematic "shift towards zero" as each group of customers slows down its collective level of purchasing over time. This is clear evidence that a substantial degree of customer dropout is taking place, and therefore confirms the need for the two different
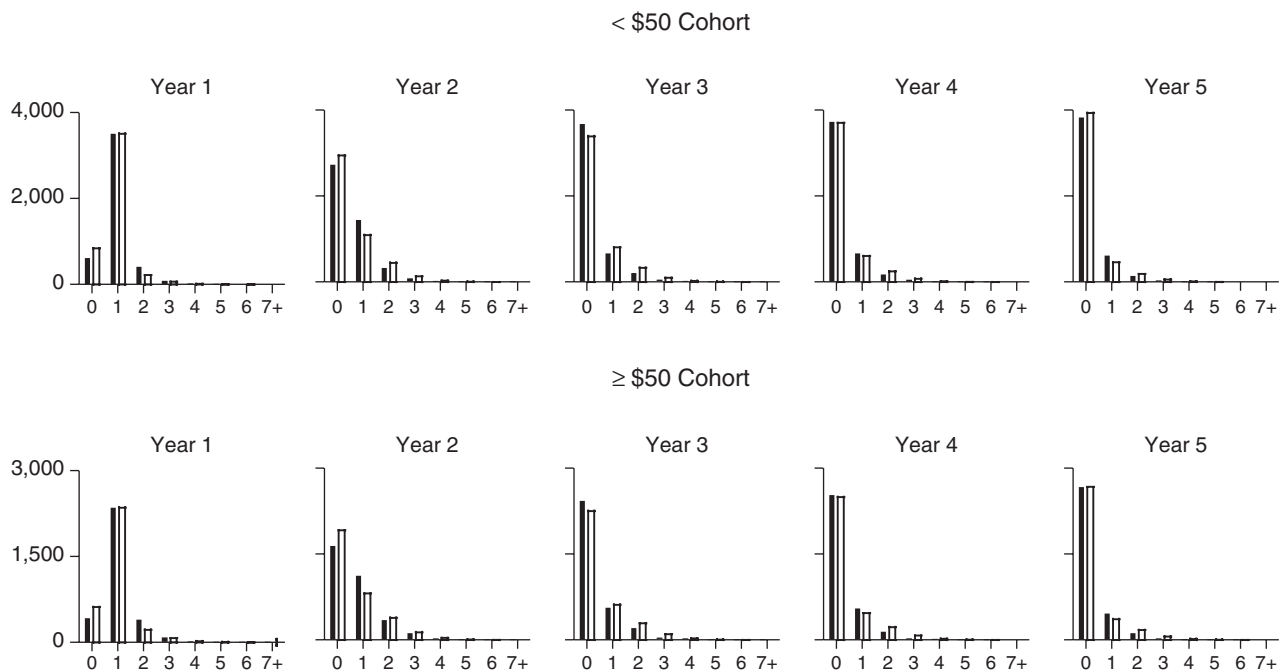
< $50 Cohort

≥ $50 Cohort

FIGURE 2

Comparing the Actual (Solid Bar) and Fitted (Clear Bar) Distributions of the Number of Transactions per Year, by Cohort

TABLE 3	Annual and Combined 5-Year MAPE Between the Actual and Fitted Values for Each Cohort					
COHORT	**YEAR 1**	**YEAR 2**	**YEAR 3**	**YEAR 4**	**YEAR 5**	**COMBINED**
<$50	10.1%	15.3%	12.9%	3.7%	7.2%	9.8%
≥$50	12.1%	19.2%	10.9%	7.1%	6.1%	11.1%

behavioral processes at the heart of the Pareto/NBD model.

Another way of summarizing model fit is to compare the model-based estimate of the average annual number of transactions per customer ($E[X(t, t + 1)]$) with the observed averages (as computed using the data in Table 1). Defining the random variable $X(t)$ as the number of transactions occurring in the interval $(0, t]$, we know from Schmittlein et al. (1987) that

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

Clearly $E[X(t, t + 1)] = E[X(t + 1)] - E[X(t)]$, so

$$E[X(t, t + 1)|r, \alpha, s, \beta]$$

$$= \frac{r\beta}{\alpha(s - 1)}\left[\left(\frac{\beta}{\beta + t}\right)^{s-1} - \left(\frac{\beta}{\beta + t + 1}\right)^{s-1}\right]. \quad (9)$$

For the five-parameter model (i.e., the basic Pareto/NBD model augmented with a "spike" at $x = 1$ for the first year), we have

$$E[X(t, t + 1)|r, \alpha, s, \beta, \pi] =$$

$$\begin{cases} \pi + (1 - \pi)\dfrac{r\beta}{\alpha(s - 1)}\left[1 - \left(\dfrac{\beta}{\beta + 1}\right)^{s-1}\right] & \text{if } t = 0 \\ \dfrac{r\beta}{\alpha(s - 1)}\left[\left(\dfrac{\beta}{\beta + 1}\right)^{s-1} - \left(\dfrac{\beta}{\beta + t + 1}\right)^{s-1}\right] & \text{if } t > 0. \end{cases}$$

$$(10)$$

The correspondence between the model-based estimates of the expected number of transactions per customer per year and the actual annual averages is examined in Figure 3. While there are some modest deviations, the overall fit is good. Furthermore, these annual deviations tend to cancel out. For the $< \$50$ cohort, the actual average number of transactions across the 5 years is 2.39, while the model estimate is 2.40; for the $\geq \$50$ cohort, both the actual and predicted average number of transactions across the 5 years is 2.80.

One additional validation exercise is to determine how robust the model is when we limit the number of histograms used to estimate it. Such a task also serves as a type of "holdout test" to see if it is appropriate to project the behavioral patterns beyond the observed 5-year period. Instead of using all 5 years of data to estimate the model, we re-estimate the model using only the first 3 years of data. We wish to see how well we can predict the histograms for years 4 and 5 despite the fact that no data from those years were used for parameter estimation. Figure 4 offers a comparison of the model predictions for this limited specification and the actual values for years 4 and 5. The close correspondence of these histograms provides strong evidence of the model's capabilities. The mean absolute percentage errors for the predictions—individually and across both years (combined)—are given in Table 4; we note that they are only slightly worse than those numbers obtained when the years 4 and 5 data were used to estimate the model parameters (Table 3). This is a tough test for any model, especially one that is calibrated on such a limited amount of data.

Having established the validity of our modeling approach, we now turn to the question that motivated the Mason (2003) case in the first place.

WHAT IS A NEW *TUSCAN LIFESTYLES* CUSTOMER WORTH?

The Pareto/NBD model enables us to predict the expected transaction stream for a new customer. However, to assess the expected lifetime value for a customer, we also need to predict the monetary amount associated with each purchase. Following Fader, Hardie, and Lee (2005), if we can assume that the monetary value of each transaction is independent of the underlying transaction process—something we must do here, given the nature of the data given in the *Tuscan*
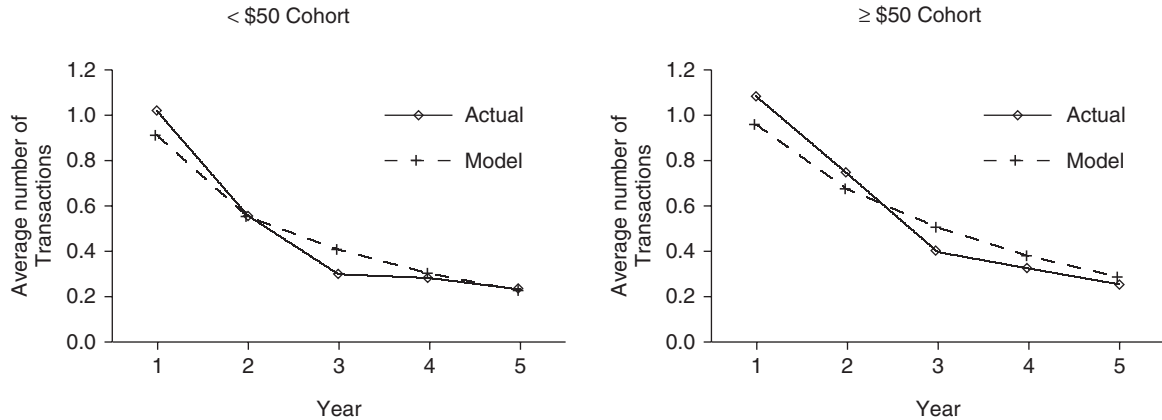
FIGURE 3

Comparing the Average Annual Number of Transactions per Customer with the Corresponding Model-Based Estimates, by Cohort
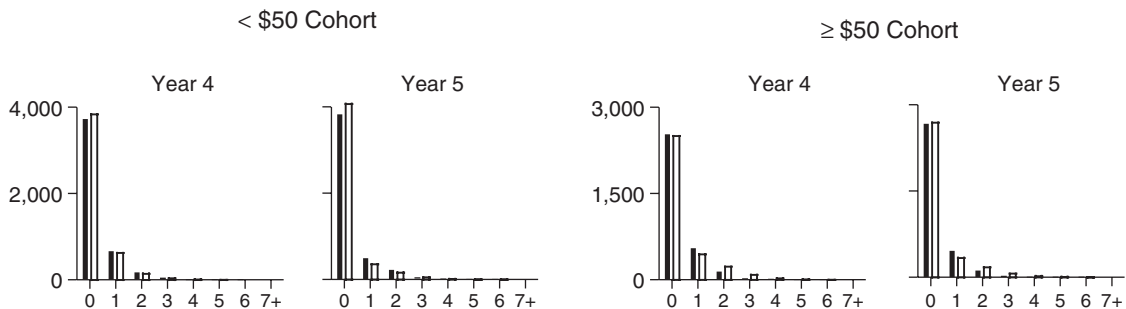
FIGURE 4

Comparing the Actual (Solid Bar) and Predicted (Clear Bar) Distributions of the Number of Transactions in Years 4 and 5 (Given Parameters Estimated Using the First Three Years of Data), by Cohort

TABLE 4	Years 4 and 5 and Combined MAPE Between the Actual and Predicted Values for Each Cohort (Given Parameters Estimated Using the First 3 Years of Data)

COHORT	YEAR 4	YEAR 5	COMBINED
< $50	4.6%	10.0%	7.3%
≥ $50	8.3%	7.6%	8.0%

Lifestyles case—the value per transaction (revenue per transaction × contribution margin) can be factored out and we can focus on forecasting the "flow" of future transactions, discounted to yield a present value. Fader, Hardie, and Lee (2005) call this quantity "discounted expected transactions," or DET; it is the effective number of repeat transactions that a customer will make, discounted back to the time of acquisition. In other words, a transaction that occurs, say, 10 years in the future, is only worth a fraction of a transaction at time zero. DET is the sum of these "fractional transactions" and therefore captures both the overall number of them as well as their spread over time. This number of discounted expected transactions can then be rescaled by a value "multiplier" to yield an overall estimate of expected lifetime value:

$$E(CLV) = \text{margin} \times E(\text{revenue/transaction}) \times DET.$$

$$(11)$$

Fader, Hardie, & Lee (2005) present an expression for DET as implied by the Pareto/NBD model. However, we cannot use it in this setting because of the

modification to the basic model to accommodate the more-than-expected number of people making just one repeat purchase in the first year. We can therefore compute DET in the following manner:

$$DET = \sum_{t=0}^{\infty} \frac{E[X(t, t + 1)]}{(1 + d)^{t+0.5}}, \qquad (12)$$

where d is the annual discount factor and $E[X(t, t + 1)]$ is computed using the expression given in (10). As the transactions can occur at any point in time during the year, we discount them as if, on average, they occur in the middle of the year. Note that this expression for DET does not include the initial transaction that signals the start of the customer's relationship with the firm. (If we wish to include this transaction—which we would need to do if we wish to set an upper bound for acquisition spend—we simply add 1 to our estimate of DET.)

Coming to the revenue/transaction component of the definition of CLV, we note that the *Tuscan Lifestyles* case only provides annual summary statistics of the spending levels for each cohort (Mason, 2003, Exhibit 3). While it is easy to conceptualize stochastic models to capture the random variation in revenue/transaction over time (Fader, Hardie, & Lee, 2005), it would be difficult to estimate reliably parameters of such a model given the limited data on monetary value available here. Since the two groups were defined on the basis of initial expenditure, this removes much of the cross-sectional variation in revenue/transaction. Thus, it is more appropriate to assume a constant level for the purchase amounts within each group of customers. The case data indicate that the mean spending level across the 5 years for the $< \$50$ group is $(32.09 + 41.78 + 51.05 + 52.43 + 53.63)/5 = \$46.20$ per transaction, while for the $\geq \$50$ group it is $(93.46 + 74.02 + 67.75 + 67.12 + 78.26)/5 = \$76.12$. Finally, we follow the case and use a fixed margin of 42% for every transaction, and a discount factor of 10% for our CLV calculations.

Using (10) and (12), we find that DET equals 2.36 for the $< \$50$ group and 2.77 for the $\geq \$50$ group. (In evaluating (12), we terminate the series at 100 years, which effectively represents infinity.) It follows that our estimate of expected CLV for the $< \$50$ group is $46, while the expected CLV for a randomly-chosen member of the $\geq \$50$ group is almost double

this value, at $89. Clearly, a customer who makes a high-value first purchase with *Tuscan Lifestyles* is more valuable in the long run compared to a customer who makes a low-value first purchase; the lone datapoint of the value of the first purchase is reasonably discriminating in determining a customer's future worth. Most of this difference is due to the fact that the average order size is 65% higher for the $\geq \$50$ cohort; in contrast, DET for the $\geq \$50$ cohort is only 17% higher than the corresponding number for the $< \$50$ cohort.

The equivalent 5-year DET numbers using the annual averages computed using the data in Table 1 are 2.04 and 2.39, resulting in "5-year lifetime value" estimates of $40 and $76, respectively. Because of the truncation at the end of 5 years, these numbers underestimate the true expected lifetime value by 14%.

Some analysts may be willing to live with a 14% error for the sake of analytical simplicity. However, we cannot be sure that the underestimation will always be so low, for instance, when the variation in transaction rates and dropout rates is high. For the data at hand, not only is the mean purchasing rate low and the dropout rate high for both cohorts, but the variation in transaction rates and dropout rates for the cohorts is also quite low. In other studies (e.g., Fader, Hardie, & Lee, 2005), considerably higher heterogeneity (along with faster purchasing and slower dropout) have been observed. Thus, the 14% underestimation in this case is a very modest number; in many other settings, the impact of ignoring the future when performing CLV calculations will likely be much larger. And beyond the CLV calculation per se, the use of the model offers many other useful diagnostics as discussed earlier and below.

Referring back to Figure 1, the between-cohort differences in the distributions of the dropout rates are smaller than those for the transaction rates. While the mean $(\beta/(s - 1))$ and median $(\beta(2^{1/s} - 1))$ lifetimes are slightly higher for the $\geq \$50$ cohort (3.5 and 2.4 years versus 3.4 and 2.2 years), the differences in the survival curves (Figure 5, left side) are negligible. Thus, the differences in DET are driven by differences in the transaction rates. We note that the mean of the transaction rate distribution is 0.88 (purchases per annum while alive) for the $< \$50$ cohort and 1.04 for the $\geq \$50$ cohort. This difference is reflected in the
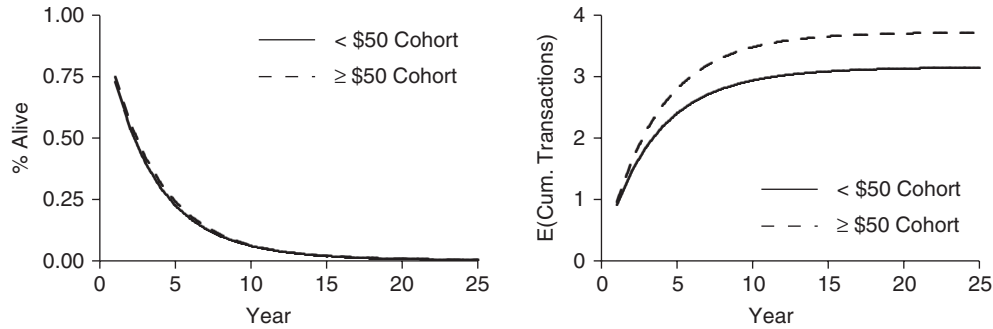
FIGURE 5

Plots of the Percentage Number of Customers Still Alive and the Expected Cumulative Number of Transactions per Customer for Years 1–25, by Cohort

plots of expected cumulative transactions (undiscounted), given on the right side of Figure 5.

As a final illustration of the value-added associated with the use of a stochastic model of buyer behavior, let us consider the question of variability in CLV (or DET). To explore this, we simulate purchase sequences for each customer, which are then discounted to give "discounted transaction" numbers. The between-customer distribution of this quantity is reported in Figure 6 for both cohorts. This figure shows how the discounted transactions are spread around the expected DET for each cohort; computing the average of these numbers yields the average DET for each cohort, as reported above. We note that while the variance in transaction rates is lower for the ≥ $50 cohort (Figure 1), the

variance in the discounted number of transactions is actually higher for this cohort (2.67 versus 2.14 for the ≥ $50 cohort).

If we had sufficient data to estimate a stochastic model for revenue/transaction, we could augment our estimates of expected CLV by the full distribution of CLV across the customer base (and associated summary statistics).

DISCUSSION AND CONCLUSIONS

The *Tuscan Lifestyles* case offers a simple new twist on the standard view of how to value a newly acquired customer, highlighting how standard retention-based
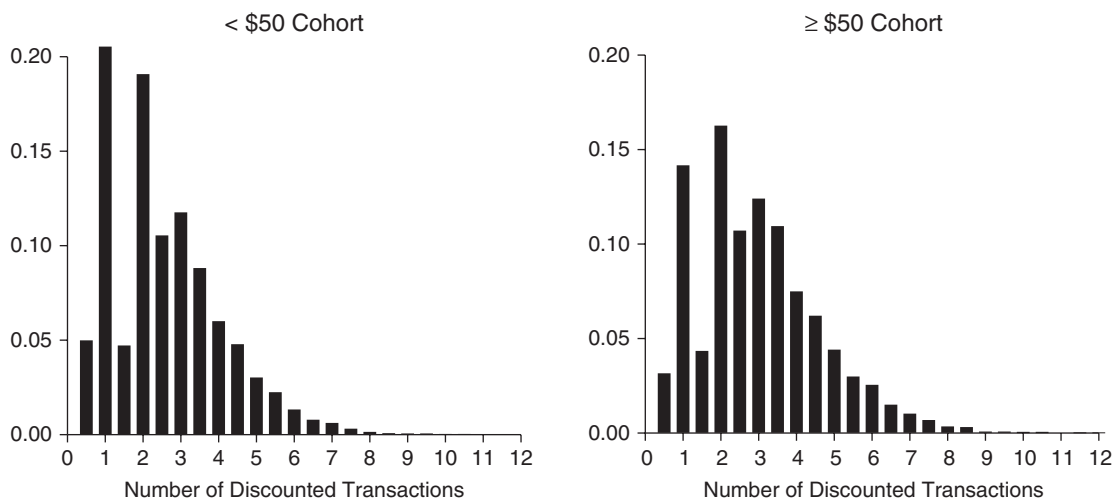
FIGURE 6

Distribution of Discounted Number of Transactions per Customer, by Cohort

approaches to the calculation of expected CLV are impractical in a noncontractual setting. It is a simple exercise to use the data presented in the case to arrive at an estimate of "expected 5-year CLV." However, if we wish to arrive at an estimate that includes the customer's "life" beyond 5 years or are interested in, say, sorting out the purchasing process (while alive) from the attrition process or computing the distribution of CLV, we need to use a formal model of buying behavior. While the Pareto/NBD model is a natural starting point, existing results do not allow us to estimate the model parameters using the data summaries presented in the case. A key contribution of this paper is the derivation of an expression that enables us to do this.

Our estimated parameters and subsequent calculations offer useful insights that could not have been obtained without the formal model. For instance, we were able to decompose the expected CLV into four factors, namely, purchasing while active, dropout, surge in sales in the first year, and monetary value of the average purchase. We observed a kind of "triple jeopardy" in that the more valuable cohort proved to be better on the three most critical factors (i.e., all but the first-year sales surge). This observation by itself deserves additional study, and may be the basis for an interesting "empirical generalization" about CLV differences across groups. By simply eye-balling the raw data, it might be possible to identify the existence of these factors, but it is impossible to assess their magnitudes, and, more importantly, the difference in their magnitudes across the two cohorts. For example, one can observe a considerable dropout rate in both cohorts, but cannot ascertain how the within-cohort distributions for the dropout rates might be different. Similarly, a spike in purchases in the first year is quite evident from the histograms, but without the underlying "organic" model of purchase, the magnitude of the spike cannot be obtained.

It is easy to see how these insights and projections can be of use to the management of *Tuscan Lifestyles* (and many other firms that face similar issues). Besides being able to judge the economic efficiency of different kinds of acquisition strategies, the model presented here can help managers determine better ways to define cohorts—does it make the most sense to divide customers on the basis of initial expenditure, or would

other kinds of splits yield more dramatic differences between groups of customers? These differences should be gauged not only in terms of overall expected CLV for each group but also in terms of the Pareto/NBD model components. Maybe a certain kind of split can lead to a greater degree of homogeneity in each group's transaction rates and/or dropout rates, thereby reducing some uncertainty about their future behavior and making it easier to target members of each group. There are clearly many substantive benefits that arise from this kind of analysis.

From a methodological standpoint, the move from detailed transaction data to histograms raises other questions as well. How about data structures that lie somewhere in between these two extremes? For instance, it is easy to imagine firms maintaining "interval-censored" data, that is, period-by-period counts for each customer. Some ideas about how to develop models using this kind of data structure are explored by Fader and Hardie (2005). Other questions relate to the length of the "window" for the censoring process (e.g., quarterly histograms versus yearly histograms) and the number of histograms needed to obtain stable parameter estimates. All in all, there are many promising research opportunities to be pursued down this path.

Although these methodological questions may be straying pretty far from the original issues raised in the *Tuscan Lifestyles* case, they provide proof of the healthy links that exist between well-formed managerial questions and appropriately constructed empirical models. New developments in one area frequently open up new possibilities in another area, to the benefit of everyone on both sides. We see *Tuscan Lifestyles* as the beginning of such a dialogue, and we look forward to continuing the conversation.

REFERENCES

Abramowitz, M., & Stegun, I. A. (Eds.). (1972). Handbook of Mathematical Functions. New York: Dover Publications.

Agrawal, R., & Srikant, R. (2000). Privacy-Preserving Data Mining. Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data. Dallas, Texas, May 15–18 (pp. 439–450). New York: ACM.

Berger, P. D., Weinberg, B., & Hanna, R. C. (2003). Customer Lifetime Value Determination and Strategic Implications for a Cruise-Ship Company. Journal of Database

Marketing & Customer Strategy Management, 11, 40–52.

Carey, P. (2004). Data Protection: A Practical Guide to UK and EU Law (2nd ed.). Oxford, UK: Oxford University Press.

Casella, G., & Berger, R. L. (2002). Statistical Inference (2nd ed.). Pacific Grove, CA: Duxbury.

Fader, P. S., & Hardie, B. G. S. (2002). A Note on an Integrated Model of Customer Buying Behavior. European Journal of Operational Research, 139, 682–687.

Fader, P. S., & Hardie, B. G. S. (2005). Implementing the Pareto/NBD Model Given Interval-Censored Data. Retrieved April 28, 2007, from http://brucehardie.com/notes/011/

Fader, P. S., & Hardie, B. G. S. (2006). Deriving an Expression for $P(X(t) = x)$ Under the Pareto/NBD Model. Retrieved April 28, 2007, from http://brucehardie.com/notes/012/

Fader, P. S., Hardie, B. G. S., & Lee, K. L. (2005). RFM and CLV: Using Iso-value Curves for Customer Base Analysis. Journal of Marketing Research, 42, 415–430.

Mielikäinen, T. (2004). Privacy Problems with Anonymized Transaction Databases. In S. Arikawa & E. Suzuki (Eds.), Discovery Science: Proceedings of the 7th International Conference (DS2004). Lecture Notes in Computer Science, 3245 (pp. 219–229). Berlin: Springer.

Mason, C. H. (2003). Tuscan Lifestyles: Assessing Customer Lifetime Value. Journal of Interactive Marketing, 17, 54–60.

Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). Introduction to the Theory of Statistics, (3rd ed.). New York: McGraw-Hill Publishing Company.

Schmittlein, D. C, Morrison, D. G., & Colombo, R. (1987). Counting Your Customers: Who Are They and What Will They Do Next? Management Science, 33, 1–24.

Schmittlein, D. C., & Peterson, R. A. (1994). Customer Base Analysis: An Industrial Purchase Process Application. Marketing Science, 13, 41–67.

Singleton, S. (2004). Tolley's Data Protection Handbook (3rd ed.). Croyden, Surrey: LexisNexis UK.

TECHNICAL APPENDIX

Schmittlein, Morrison, and Colombo (1987) and Fader and Hardie (2006) derive expressions for $P(X(t) = x)$, where the random variable $X(t)$ denotes the number of transactions observed in the time interval $(0, t]$, as implied by the Pareto/NBD model assumptions. In this appendix, we derive the corresponding expression for $P(X(t, t + 1) = x)$, where the random variable $X(t, t + 1)$ denotes the number of transactions observed in the time interval $(t, t + 1]$.

Let us first review the assumptions underlying the Pareto/NBD model:

i. Customers go through two stages in their "lifetime" with a specific firm: they are "alive" for some period of time, then become permanently inactive.

ii. While alive, the number of transactions made by a customer follows a Poisson process with transaction rate λ. This implies that the probability of observing x transactions in the time interval $(0, t]$ is given by

$$P(X(t) = x|\lambda) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}, \quad x = 0, 1, 2, \ldots .$$

It also implies that, assuming the customer is alive through the time interval $(t_a, t_b]$,

$$P(X(t_a, t_b) = x|\lambda) = \frac{[\lambda(t_b - t_a)]^x e^{-\lambda(t_b - t_a)}}{x!}, \quad x = 0, 1, 2, \ldots .$$

iii. A customer's unobserved "lifetime" of length ω (after which he is viewed as being inactive) is exponentially distributed with dropout rate μ:

$$f(\omega|\mu) = \mu e^{-\mu\omega}.$$

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

$$g(\lambda|r, \alpha) = \frac{\alpha^r \lambda^{r-1} e^{-\lambda\alpha}}{\Gamma(r)}. \tag{A1}$$

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

$$g(\mu|s, \beta) = \frac{\beta^s \mu^{s-1} e^{-\mu\beta}}{\Gamma(s)}. \tag{A2}$$

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

Suppose we know an individual's unobserved latent characteristics λ and μ. For $x > 0$, there are two ways x purchases could have occurred in the interval $(t, t + 1]$:

i. The individual was alive at t and remained alive through the whole interval; this occurs with probability $e^{-\mu(t+1)}$. The probability of the individual making x purchases, given that he was alive during the whole interval, is $\lambda^x e^{-\lambda}/x!$. It follows that the probability of remaining alive through the interval $(t, t + 1]$ and making x purchases is

$$\frac{\lambda^x e^{-\lambda x} e^{-\mu(t+1)}}{x!}. \tag{A3}$$

ii. The individual was alive at t but "died" at some point ω $(< t + 1)$, making x purchases in the interval $(t, \omega]$. The probability of this occurring is

$$\int_{t}^{t+1} \frac{[\lambda(\omega - t)]^x e^{-\lambda(\omega - t)}}{x!} \mu e^{-\mu\omega} d\omega = e^{-\mu t} \frac{\lambda^x \mu}{(\lambda + \mu)^{x+1}} \int_{0}^{1} \frac{(\lambda + \mu)^{x+1} s^x e^{-(\lambda + \mu)s}}{x!} ds,$$

which, noting that the integrand is an Erlang-$(x + 1)$ pdf, equals

$$e^{-\mu t} \left(\frac{\lambda}{\lambda + \mu}\right)^x \left(\frac{\mu}{\lambda + \mu}\right) \left[1 - e^{-(\lambda + \mu)} \sum_{i=0}^{x} \frac{[(\lambda + \mu)]^i}{i!}\right]. \tag{A4}$$

These two scenarios also hold for the case of $x = 0$ but need to be augmented by an additional reason as to why no purchases could have occurred in the interval $(t, t + 1]$: The individual was dead at the beginning of the interval, which occurs with probability

$$1 - e^{-\mu t}. \tag{A5}$$

Combining (A3)–(A5) gives us the following expression for the probability of observing x purchases in the interval $(t, t + 1]$, conditional on λ and μ:

$$P(X(t, t + 1) = x | \lambda, \mu) = \delta_{x=0}\left[1 - e^{-ut}\right] + \frac{\lambda^x e^{-\lambda} e^{-\mu(t+1)}}{x!} + \left(\frac{\lambda}{\lambda + \mu}\right)^x \left(\frac{\mu}{\lambda + \mu}\right) e^{-\mu t}$$

$$- \left(\frac{\lambda}{\lambda + \mu}\right)^x \left(\frac{\mu}{\lambda + \mu}\right) e^{-\lambda} e^{-\mu(t+1)} \sum_{i=0}^{x} \frac{[(\lambda + \mu)]^i}{i!}. \tag{A6}$$

In reality, we never know an individual's latent characteristics; we therefore remove the conditioning on λ and μ by taking the expectation of (A6) over the distributions of Λ and M:

$$P(X(t, t + 1) = x | r, \alpha, s, \beta) = \int_{0}^{\infty} \int_{0}^{\infty} P(X(t, t + 1) = x | \lambda, \mu) g(\lambda | r, \alpha) g(\mu | s, \beta) d\lambda \, d\mu. \tag{A7}$$

Substituting (A1), (A2), and (A6) in (A7) gives us

$$P(X(t, t + 1) = x | r, \alpha, s, \beta) = \delta_{x=0} A_1 + A_2 + A_3 - \sum_{i=0}^{x} \frac{1}{i!} A_4 \tag{A8}$$

where

$$A_1 = \int_{0}^{\infty} [1 - e^{-\mu t}] g(\mu | s, \beta) \, d\mu \tag{A9}$$

$$A_2 = \int_{0}^{\infty} \int_{0}^{\infty} \frac{\lambda^x e^{-\lambda} e^{-\mu(t+1)}}{x!} g(\lambda | r, \alpha) g(\mu | s, \beta) d\lambda \, d\mu \tag{A10}$$

$$A_3 = \int_{0}^{\infty} \int_{0}^{\infty} \left(\frac{\lambda}{\lambda + \mu}\right)^x \left(\frac{\mu}{\lambda + \mu}\right) e^{-\mu t} g(\lambda | r, \alpha) g(\mu | s, \beta) d\lambda \, d\mu \tag{A11}$$

$$A_4 = \int_{0}^{\infty} \int_{0}^{\infty} \left(\frac{\lambda}{\lambda + \mu}\right)^x \left(\frac{\mu}{\lambda + \mu}\right) (\lambda + \mu)^i e^{-\mu(t+1)} g(\lambda | r, \alpha) g(\mu | s, \beta) d\lambda \, d\mu \tag{A12}$$

Solving (A9) and (A10) is trivial:

$$A_1 = 1 - \left(\frac{\beta}{\beta + t}\right)^s \tag{A13}$$

$$A_2 = \frac{\Gamma(r + x)}{\Gamma(r)x!}\left(\frac{\alpha}{\alpha + 1}\right)^r\left(\frac{1}{\alpha + 1}\right)^x\left(\frac{\beta}{\beta + t + 1}\right)^s \tag{A14}$$

To solve (A11), consider the transformation $Y = M/(\Lambda + M)$ and $Z = \Lambda + M$. Using the transformation technique (Casella & Berger, 2002, Section 4.3, pp. 156–162; Mood, Graybill, & Boes, 1974, Section 6.2, p. 204ff), it follows that the joint distribution of Y and Z is

$$g(y, z \mid \alpha, \beta, r, s) = \frac{\alpha^r \beta^s}{\Gamma(r)\Gamma(s)}y^{s-1}(1 - y)^{r-1}z^{r+s-1}e^{-z(\alpha - (\alpha - \beta)y)}. \tag{A15}$$

Noting that the inverse of this transformation is $\lambda = (1 - y)z$ and $\mu = yz$, it follows that

$$\begin{aligned}
A_3 &= \int_0^1\int_0^\infty y(1 - y)^x e^{-yzt}g(y,z \mid \alpha, \beta, r, s)\, dz\, dy \\
&= \frac{\alpha^r \beta^s}{\Gamma(r)\Gamma(s)}\int_0^1\int_0^\infty y^s(1 - y)^{r+x-1}z^{r+s-1}e^{-z(\alpha - (\alpha - (\beta + t))y)}dz\, dy \\
&= \frac{1}{B(r, s)}\frac{\alpha^r \beta^s}{\alpha^{r+s}}\int_0^1 y^s(1 - y)^{r+x-1}\left[1 - \left(\frac{\alpha - (\beta + t)}{\alpha}\right)y\right]^{-(r+s)}dy,
\end{aligned}$$

which, recalling Euler's integral for the Gaussian hypergeometric function,[1] equals

$$\left(\frac{\beta}{\alpha}\right)^s\frac{B(r + x, s + 1)}{B(r, s)}{}_2F_1\left(r + s, s + 1; r + s + x + 1; \frac{\alpha - (\beta + t)}{\alpha}\right). \tag{A16}$$

Looking closely at (A16), we see that the argument of the Gaussian hypergeometric function, $\frac{\alpha - (\beta + t)}{\alpha}$, is guaranteed to be bounded between 0 and 1 when $\alpha \geq \beta + t$, thus ensuring convergence of the series representation of the function. However, when $\alpha < \beta + t$ we can be faced with the situation where $\frac{\alpha - (\beta + t)}{\alpha} < -1$, in which case the series is divergent.

Applying the linear transformation (Abramowitz & Stegun, 1972, equation 15.3.4)

$$_2F_1(a, b; c; z) = (1 - z)^{-a}\,_2F_1\left(a, c - b; c; \frac{z}{z - 1}\right), \tag{A17}$$

gives us

$$A_3 = \frac{\alpha^r \beta^s}{(\beta + t)^{r+s}}\frac{B(r + x, s + 1)}{B(r, s)}\,_2F_1\left(r + s, r + x; r + s + x + 1; \frac{\beta + t - \alpha}{\beta + t}\right). \tag{A18}$$

We note that the argument of the above Gaussian hypergeometric function is bounded between 0 and 1 when $\alpha \leq \beta + t$. We therefore present (A16) and (A18) as solutions to (A11), using (A16) when $\alpha \geq \beta + t$ and (A18) when $\alpha \leq \beta + t$. We can write this as

$$A_3 = \alpha^r \beta^s\frac{B(r + x, s + 1)}{B(r, s)}B_1 \tag{A19}$$

---

[1] See http://functions.wolfram.com/07.23.07.0001.01

where

$$
B_1 = \begin{cases} {}_2F_1\!\left(r+s, s+1; r+s+x+1; \dfrac{\alpha-(\beta+t)}{\alpha}\right)\!\Big/\alpha^{r+s} & \text{if } \alpha \geq \beta+t \\[2ex] {}_2F_1\!\left(r+s, r+x; r+s+x+1; \dfrac{\beta+t-\alpha}{\beta+t}\right)\!\Big/(\beta+t)^{r+s} & \text{if } \alpha \leq \beta+t. \end{cases} \tag{A20}
$$

To solve (A12), we also make use of the transformation $Y = M/(\Lambda + M)$ and $Z = \Lambda + M$. Given (A15), it follows that

$$
\begin{aligned}
A_4 &= \int_0^1\!\int_0^\infty y(1-y)^x z^i e^{-(1-y)z} e^{-yz(t+1)} g(y,z\,|\,\alpha,\beta,r,s)\,dz\,dy \\[1ex]
&= \frac{\alpha^r\beta^s}{\Gamma(r)\Gamma(s)}\int_0^1\!\int_0^\infty y^s(1-y)^{r+x-1} z^{r+s+x-1} e^{-z(\alpha+1-(\alpha-(\beta+t))y)}\,dz\,dy \\[1ex]
&= \frac{\Gamma(r+s+i)}{\Gamma(r)\Gamma(s)}\frac{\alpha^r\beta^s}{(\alpha+1)^{r+s+i}}\int_0^1 y^s(1-y)^{r+x-1}\left[1-\left(\frac{\alpha-(\beta+t)}{\alpha+1}\right)y\right]^{-(r+s+i)}dy,
\end{aligned}
$$

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

$$
\frac{\Gamma(r+s+i)}{\Gamma(r+s)}\frac{\alpha^r\beta^s}{(\alpha+1)^{r+s+i}}\frac{B(r+x,s+1)}{B(r,s)}\,{}_2F_1\!\left(r+s+i, s+1; r+s+x+1; \frac{\alpha-(\beta+t)}{\alpha+1}\right). \tag{A21}
$$

Noting that the argument of the Gaussian hypergeometric function is only guaranteed to be bounded between 0 and 1 when $\alpha \geq \beta + t$, we apply the linear transformation (A17), which gives us

$$
A_4 = \frac{\Gamma(r+s+i)}{\Gamma(r+s)}\frac{\alpha^r\beta^s}{(\beta+t+1)^{r+s+i}}\frac{B(r+x,s+1)}{B(r,s)}\,{}_2F_1\!\left(r+s+i, r+x; r+s+x+1; \frac{\beta+t-\alpha}{\beta+t+1}\right). \tag{A22}
$$

The argument of the above Gaussian hypergeometric function is bounded between 0 and 1 when $\alpha \leq \beta + t$. We therefore present (A21) and (A22) as solutions to (A12): We use (A21) when $\alpha \geq \beta + t$ and (A22) when $\alpha \leq \beta + t$. We can write this as

$$
A_4 = \alpha^r\beta^s\frac{\Gamma(r+s+i)}{\Gamma(r+s)}\frac{B(r+x,s+1)}{B(r,s)}B_2 \tag{A23}
$$

where

$$
B_2 = \begin{cases} {}_2F_1(r+s+i, s+1; r+s+x+1; \frac{\alpha-(\beta+t)}{\alpha+1})/(\alpha+1)^{r+s+i} & \text{if } \alpha \geq \beta+t \\[1ex] {}_2F_1(r+s+i, r+x; r+s+x+1; \frac{\beta+t-\alpha}{\beta+t+1})/(\beta+t+1)^{r+s+i} & \text{if } \alpha \leq \beta+t. \end{cases} \tag{A24}
$$

Substituting (A13), (A14), (A19), and (A23) in (A8) gives us the following expression for the probability of observing x transactions in the time interval $(t, t+1)$:

$$
\begin{aligned}
P(X(t, t+1) = x\,|\,r, \alpha, s, \beta) &= \delta_{x=0}\left[1-\left(\frac{\beta}{\beta+t}\right)^s\right] + \frac{\Gamma(r+x)}{\Gamma(r)x!}\left(\frac{\alpha}{\alpha+1}\right)^r\times\left(\frac{1}{\alpha+1}\right)^x\!\left(\frac{\beta}{\beta+t+1}\right)^s \\[1ex]
&\quad + \alpha^\alpha\beta^s\frac{B(r+x,s+1)}{B(r,s)}\left\{B_1 - \sum_{i=0}^x\frac{\Gamma(r+s+i)}{\Gamma(r+s)}\frac{1}{i!}B_2\right\}
\end{aligned} \tag{A25}
$$

where expressions for B_1 and B_2 are given in (A20) and (A24), respectively.