Estimating CLV Using Aggregated Data: The *Tuscan Lifestyles* Case Revisited

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

September 2006

†Peter S. Fader is the Frances and Pei-Yuan Chia Professor of Marketing at the Wharton School of the University of Pennsylvania (address: 749 Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104-6340; phone: 215.898.1132; email: faderp@wharton.upenn.edu; web: www.petefader.com). Bruce G. S. Hardie is Associate Professor of Marketing, London Business School (email: bhardie@london.edu; web: www.brucehardie.com). Kinshuk Jerath is a doctoral candidate at the Wharton School of the University of Pennsylvania (email: kinshuk@wharton.upenn.edu). The second author acknowledges the support of the London Business School Centre for Marketing.

Abstract

Estimating CLV Using Aggregated Data: The *Tuscan Lifestyles* Case Revisited

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 CLV are useless in a noncontractual setting. Using the data presented in the case, it is a simple exercise to compute an estimate of "expected five-year CLV". However, 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.

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.

Keywords: customer lifetime value, Pareto/NBD.

1 Introduction

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

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

(where m is the net cash-flow per period, r is the retention rate, d is the discount rate, and T is the time horizon for calculation) and claim that it can be a useful starting point for the calculation of lifetime value.

However, such an expression is useless 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).) As the point in time 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.

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 five-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).

Using these data, we can easily arrive at an estimate of "expected five-year CLV". But what about the customer's "life" beyond five 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? We feel that any serious examination of CLV

< $50 Cohort

# Orders	Year 1	2	3	4	5
0	611	2739	3687	3730	3837
1	3508	1441	671	661	626
2	416	332	207	185	141
3	94	100	68	56	38
4	21	30	19	14	10
5	7	9	1	9	3
6		3	2	2	
7		1	1		
8		1			2
9		1			
10			1		
11					
12					
13					

≥ $50 Cohort

# Orders	Year 1	2	3	4	5
0	421	1643	2430	2535	2673
1	2354	1120	562	548	463
2	397	354	214	151	120
3	91	121	53	37	30
4	20	37	27	20	7
5	6	12	6	3	2
6	5	5	2	2	1
7	1	3	1		
8	1				
9			1		
10					
11					
12					
13		1			

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.

should consider such questions. Unfortunately they cannot be answered using these data alone. This situation is not unique; other researchers (e.g., Berger et al. 2003) have performed similar analyses and therefore face similar limitations.

Therefore, 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 interrelated) 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 individual-customer level, 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 et al. 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 Section 2, we derive an expression that enables us to estimate the model parameters using such data. We then fit the model to the data (Section 3) and examine the question "what is a new *Tuscan Lifestyles* customer worth?" (Section 4).

2 Model Development

The Pareto/NBD model 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

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.

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 transactions 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 one; 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 1441 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:

$$\begin{aligned} P(X(t,t+1) = x \,|\, \lambda, \mu) = \delta_{x=0}\Big[1 - e^{-\mu t}\Big] &+ \frac{\lambda^x e^{-\lambda} e^{-\mu(t+1)}}{x!} \\ &+ \Big(\frac{\lambda}{\lambda+\mu}\Big)^x \Big(\frac{\mu}{\lambda+\mu}\Big) e^{-\mu t} \\ &- \Big(\frac{\lambda}{\lambda+\mu}\Big)^x \Big(\frac{\mu}{\lambda+\mu}\Big) e^{-\lambda} e^{-\mu(t+1)} \sum_{i=0}^{x} \frac{[(\lambda+\mu)]^i}{i!}\,, \end{aligned} \tag{2}$$

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:

$$\begin{aligned} &P(X(t,t+1) = x \,|\, r, \alpha, s, \beta) \\ &\qquad = \delta_{x=0}\left[1 - \Big(\frac{\beta}{\beta+t}\Big)^s\right] + \frac{\Gamma(r+x)}{\Gamma(r)x!}\Big(\frac{\alpha}{\alpha+1}\Big)^r \Big(\frac{1}{\alpha+1}\Big)^x \Big(\frac{\beta}{\beta+t+1}\Big)^s \\ &\qquad\quad + \alpha^r \beta^s \frac{B(r+x, s+1)}{B(r,s)} \left\{ \mathsf{B}_1 - \sum_{i=0}^{x} \frac{1}{iB(r+s,i)} \mathsf{B}_2 \right\} \end{aligned} \tag{3}$$

where

$$\mathsf{B}_1 = \begin{cases} {}_2F_1\big(r+s, s+1; r+s+x+1; \frac{\alpha-(\beta+t)}{\alpha}\big)/\alpha^{r+s} & \text{if } \alpha \geq \beta + t \\ {}_2F_1\big(r+s, r+x; r+s+x+1; \frac{\beta+t-\alpha}{\beta+t}\big)/(\beta+t)^{r+s} & \text{if } \alpha \leq \beta + t \end{cases} \tag{4}$$

$$\mathsf{B}_2 = \begin{cases} {}_2F_1\big(r+s+i, s+1; r+s+x+1; \frac{\alpha-(\beta+t)}{\alpha+1}\big)/(\alpha+1)^{r+s+i} & \text{if } \alpha \geq \beta + t \\ {}_2F_1\big(r+s+i, r+x; r+s+x+1; \frac{\beta+t-\alpha}{\beta+t+1}\big)/(\beta+t+1)^{r+s+i} & \text{if } \alpha \leq \beta + t \end{cases} \tag{5}$$

and ${}_2F_1(\cdot)$ is the Gaussian hypergeometric function. (See Fader et al. (2006) for a comprehensive technical appendix that provides a step-by-step derivation of (2)–(5).)

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, $T = 5$, $n_{0,1} = 611$, $n_{1,2} = 1441$, and so on.) The sample log-likelihood function is given by

$$LL(r, \alpha, s, \beta) = \sum_{t=0}^{T-1} \sum_{x=0}^{\infty} n_{x,t} \ln \left[P(X(t, t+1) = x \,|\, r, \alpha, s, \beta) \right] . \tag{6}$$

This can be maximized using standard numerical optimization routines.

3 Model Estimation Results

We first applied the basic model to the data in Table 1. Using (6), 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 five 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 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 purchase do not show any sharp variations. While the Pareto/NBD model is very flexible, it is not sufficiently flexible to capture this year 1 aberration — not only did it miss the spike at $x = 1$ in year one, 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 buyers that would expire one 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

$$P(X(t,t+1)=x\,|\,r,\alpha,s,\beta,\pi)$$
$$=\begin{cases}\pi+(1-\pi)P_{PNBD}(X(t,t+1)=x) & \text{if } t=0 \text{ and } x=1\\ (1-\pi)P_{PNBD}(X(t,t+1)=x) & \text{if } t=0 \text{ and } x\neq 1\\ P_{PNBD}(X(t,t+1)=x) & \text{if } t>0\end{cases} \quad (7)$$

where $P_{PNBD}(\cdot)$ is the basic Pareto/NBD probability given in (3). (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.

Cohort	r	α	s	β	π
$<$ \$50	32.83	37.21	12.13	37.74	0.63
$\geq$ \$50	148.11	142.07	29.00	98.26	0.57

Table 2: Parameter estimates by cohort

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 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.

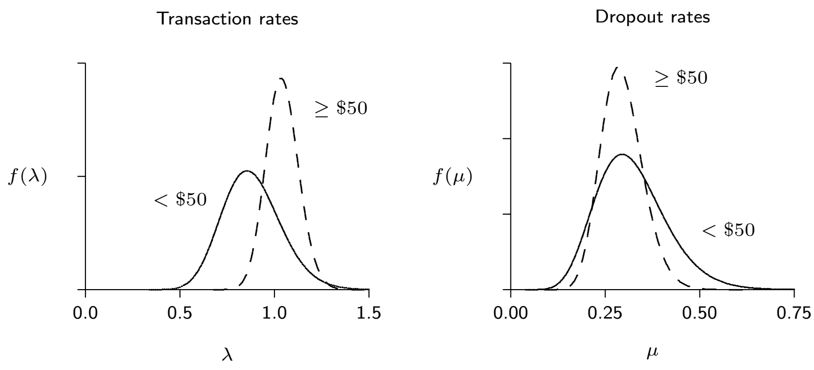

Figure 1: Heterogeneity in transaction and dropout rates for the two cohorts

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 differences in dropout tendencies both within and across each of the customer groups.

Even stronger support for the model is offered in Figure 2, which presents a side-by-side plot of the actual and fitted values for each group. It is quite remarkable to see how well a five-parameter model can capture the different shapes that can be 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 behavioral processes at the heart of the Pareto/NBD model.

Another way of summarizing model fit is to compare the predicted 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

$$E[X(t)\,|\,r,\alpha,s,\beta] = \frac{r\beta}{\alpha(s-1)}\left[1-\left(\frac{\beta}{\beta+t}\right)^{s-1}\right]. \tag{8}$$

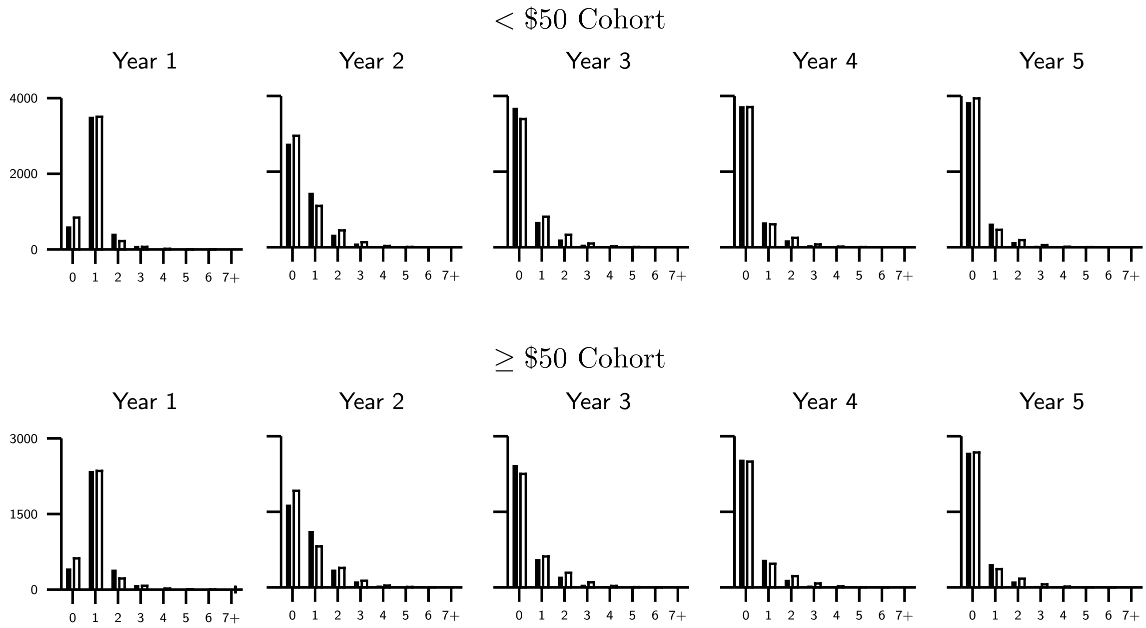

Figure 2: Comparing the the actual (solid bar) and predicted (clear bar) distributions of the number of transactions per year, by cohort

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

$$E[X(t,t+1)\,|\,r,\alpha,s,\beta] = \frac{r\beta}{\alpha(s-1)}\left[\left(\frac{\beta}{\beta+t}\right)^{s-1} - \left(\frac{\beta}{\beta+t+1}\right)^{s-1}\right]. \tag{9}$$

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

$$E[X(t,t+1)\,|\,r,\alpha,s,\beta,\pi] = \begin{cases} \pi + (1-\pi)\dfrac{r\beta}{\alpha(s-1)}\left[1 - \left(\dfrac{\beta}{\beta+1}\right)^{s-1}\right] & \text{if } t = 0 \\[2ex] \dfrac{r\beta}{\alpha(s-1)}\left[\left(\dfrac{\beta}{\beta+t}\right)^{s-1} - \left(\dfrac{\beta}{\beta+t+1}\right)^{s-1}\right] & \text{if } t > 0 \end{cases} \tag{10}$$

These annual averages are plotted 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 five years is 2.39, while the model estimate is 2.40; for the ≥ \$50 cohort, both the actual and predicted average number of transactions across the five years is 2.80.

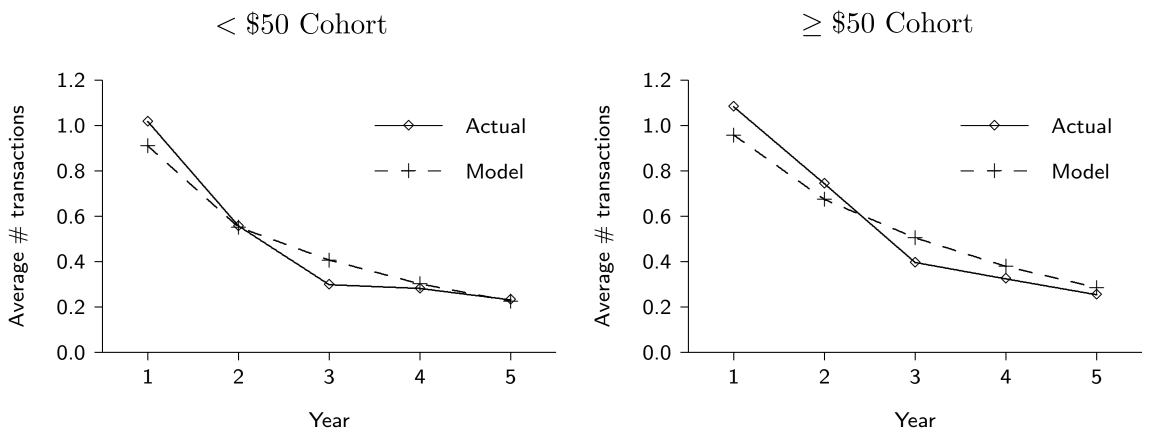

Figure 3: Plots of predicted versus actual average annual number of transactions per customer, by cohort

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 sample. Instead of using all five years of data to estimate the model, we re-estimate the model using only the first three 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. This is a tough test for any model, especially one that is calibrated on such limited amounts of data.

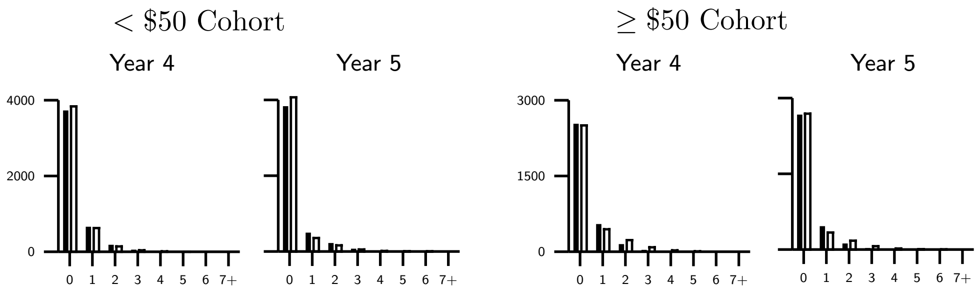

Figure 4: Plots of the predicted transaction distributions for years 4 and 5 (given parameters estimated using the first three years of data) against the actual histograms

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

4 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 et al. (2005), if we can assume that the monetary value of each transaction is independent of the underlying transaction process — something we have to do given the nature of the data given in the *Tuscan 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. We 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:

$$E(CLV) = \text{margin} \times E(\text{revenue/transaction}) \times DET \tag{11}$$

Fader et al. (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:

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

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.)

It is easy to develop stochastic models to capture the random variation in revenue/transaction over time (Fader et al. 2005), but it would be difficult to reliably estimate such a model given the limited data on monetary value in the *Tuscan Lifestyles* case. 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 data in the case (Mason 2003, Exhibit 3) indicate that the mean spending level across the five 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 data-point 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 five-year DET numbers using the annual averages computed using the data in Table 1 are 2.04 and 2.39, resulting in "five year lifetime value" estimates of \$40 and \$76, respectively. Because of the truncation at the end of five 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. More importantly, however, all we would get is an estimate of expected five-year CLV; we would have no ability to understand why we observe inter-cohort differences in DET, let alone gain any sense of the other diagnostics associated with the Pareto/NBD model.

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) for the $<$ \$50 cohort and 1.04 for the $\geq$ \$50 cohort. This difference is reflected the plots of expected cumulative transactions (undiscounted), given on the right side of Figure 5.

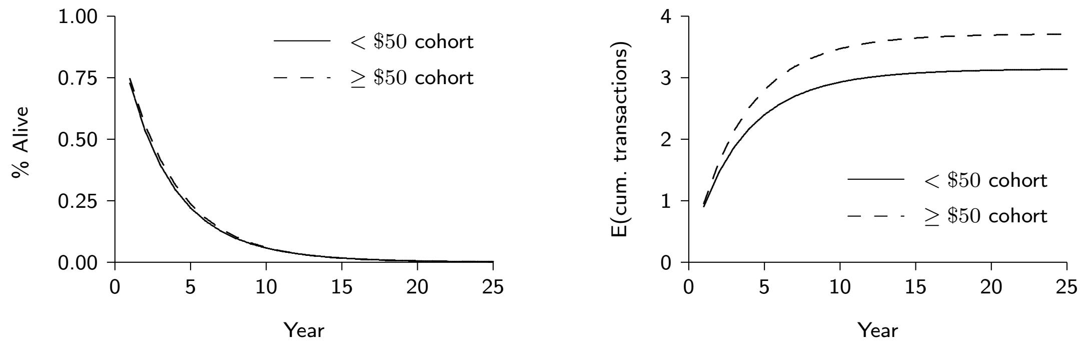

Figure 5: Plots of the percentage number of customers still alive and the expected cumulative number of transaction for years 1–25, by cohort

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; taking a simple sum 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 $\geq$ \$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 $\geq$ \$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).

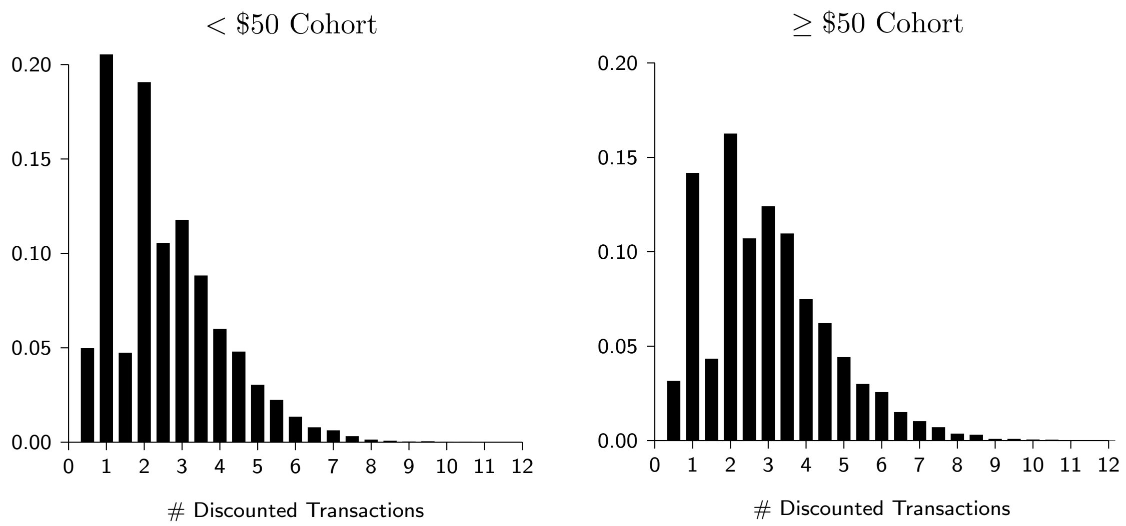

Figure 6: Distribution of discounted number of transactions per customer, by cohort

5 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 approaches to the calculation of expected CLV are useless in a noncontractual setting. It is a simple exercise to use the data presented in the case to arrive at an estimate of "expected five-year CLV". However, if we wish to arrive at an estimate 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. 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.

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, i.e., 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 vs. 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 the other 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

Berger, Paul D., Bruce Weinberg, and Richard C. Hanna (2003), "Customer Lifetime Value Determination and Strategic Implications for a Cruise-Ship Company," *Journal of Database Marketing & Customer Strategy Management*, **11** (September), 40–52.

Fader, Peter S. and Bruce G. S. Hardie (2002), "A Note on an Integrated Model of Customer Buying Behavior," *European Journal of Operational Research*, **139** (3), 682–687.

Fader, Peter S. and Bruce G. S. Hardie (2005), "Implementing the Pareto/NBD Model Given Interval-Censored Data," <`http://brucehardie.com/notes/011/`>.

Fader, Peter S., Bruce G. S. Hardie, and Kinshuk Jerath (2006), "Deriving an Expression for $P(X(t, t+\tau) = x)$ Under the Pareto/NBD Model," <`http://brucehardie.com/notes/013/`>.

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

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

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