Chapter 5
Customer Lifetime Value: Fundamentals

Abstract Customer lifetime value (LTV) is one of the cornerstones of database marketing. It is the metric by which we quantify the customer's long-term value to the firm. This chapter focuses on the fundamental methods for calculating lifetime value, centering on "simple retention models" and "migration models." We present a general approach to calculating LTV using these models, and illustrate with specific examples. We also discuss the particular case of calculating LTV when customer attrition is unobserved.

5.1 Introduction

Marketing needs to develop key metrics if it wants to become more relevant to top management. The commonly used marketing metrics are sales and market share but these measures are "dated". They are aggregate "30,000 feet" measures and do not provide the level of insight modern executives need to manage their businesses. This chapter focuses on a relatively new metric: lifetime value of a customer (LTV).

LTV has two main applications: (1) to diagnose the health of a business and (2) to assist in making tactical decisions. LTV provides a longer-run economic view of the customer and generates diagnostics based on the parameters that determine it: retention rates, sales per customer, and costs. LTV, linked with customer acquisition rates and expenditures, quantifies the long-term profitability of the firm. A firm cannot reduce customer acquisition investment without a warning signal going off: the number of newly acquired customers multiplied by their LTV would decline, indicating a long-term decline in total company profit.

LTV's tactical applications include determining how much a firm can invest to acquire customers and deciding how much service to offer a given customer. For example, a bank might decide that high-LTV customers should receive better services (e.g., a personal representative and no service fees for bank checking accounts).

This chapter covers the fundamentals of calculating LTV. Chapter 6 covers challenging issues in computing LTV and Chapter 7 provides LTV applications.

5.1.1 Definition of Lifetime Value of a Customer

The definition we will use for the lifetime value of a customer (LTV) is:

> The net present value of the profits linked to a specific customer once the customer has been acquired, after subtracting incremental costs associated with marketing, selling, production and servicing over the customer's lifetime.

There are a number of important issues implied in this definition. The firm needs to: (1) forecast future sales of a customer; (2) compute incremental costs per customer; and (3) determine the relevant discount rate to use in the present value calculation. Note also that we do not include acquisition costs as part of lifetime value. However, we often display customer acquisition cost alongside customer LTV. In this way, we gain insight on whether an unprofitable customer (for whom LTV minus acquisition cost is negative) is due to high acquisition cost or low LTV. Formally, we refer to LTV minus acquisition cost as "Customer Equity" (Blattberg et al. 2001).

5.1.2 A Simple Example of Calculating Customer Lifetime Value

Assume a firm spends $2.00 for mailing and printing a catalog which is sent to 1,000,000 prospects. The response rate to the mailing is 1%. Prospects who become customers spend $200 per year as long as they are still active customers. A customer has a probability of "attriting" ("churning")[1] each year of 20%. If a customer attrites, he or she ceases to be a customer and never returns. The firm also spends $20 per year in marketing (catalogs and service) to each active customer. The firm has a gross margin of 50% and uses a discount rate of 15%.

Table 5.1a shows the computations for the LTV of the customer just described along with the average acquisition cost per customer. We see that the acquisition cost is less than the LTV and so the firm should invest in acquiring this customer.

This example highlights some of the critical information required to compute LTV. Table 5.2 summarizes these issues and where they are covered.

---

[1] Throughout this chapter and book we will use attrite and churn interchangeably.

Table 5.1 Lifetime value and acquisition cost calculations: A simple example

Table 5.1a Lifetime value

Parameters	
Retention rate	80%
Revenues if still a customer	$200
Profit margin	50%
Gross profit if still a customer	$100
Marketing cost if still a customer	$20
Net annual profit if still a customer	$80
Discount rate	15%

Year	**Survival rate**[a]	**Expected profit**	**Discount multiplier**[b]	**Net discounted profit**
1	1.000	$80	1.000	$80
2	0.800	$64	0.870	$56
3	0.640	$51	0.756	$39
4	0.512	$41	0.658	$27
5	0.410	$33	0.572	$19
6	0.328	$26	0.497	$13
7	0.262	$21	0.432	$9
8	0.210	$17	0.376	$6
9	0.168	$13	0.327	$4
10	0.134	$11	0.284	$3
LTV = Total net discounted profit = $256				

[a]The survival rate is the probability the customer is still a customer in a given year. In this case the survival rate in year t is 0.8^{t-1}. This is because the customer has a 0.2 probability of attriting each year; hence a retention rate of 0.8, and we assume the retention rate is constant over time.
[b]Discount multiplier $= 1/(1 + \text{discount rate})^{(\text{Year}-1)}$

Table 5.1b Acquisition cost

Mail cost per prospect	$2
Number of prospect mailings	1,000,000
Total mail costs	$2,000,000
Response rate	1%
Number of customers acquired	10,000
Cost per acquired customer	**$200**

Table 5.2 Information requirements for computing LTV

Parameter	**Coverage**
Retention rates	Section 3
Unobserved attrition	Section 4
Expected revenue per customer	Section 5
Relevant costs	Chapter 6
Appropriate discount rate	Chapter 6

5.2 Mathematical Formulation of LTV

Lifetime value can be stated as:

$$LTV = \sum_{t=1}^{\infty} \frac{E[\widetilde{V}_t]}{(1+\delta)^{t-1}} \tag{5.1}$$

where:

$\widetilde{V}_t$ = a random variable representing the customer's net profit contribution during time t.
δ = the discount rate per time unit t.

Profit contribution over time is uncertain; therefore LTV is the *expected* net present value of future profit contributions. The assumptions made in quantifying these uncertain returns determine LTV. For simplicity we do not include a customer subscript in Equation 5.1. Ideally, the calculation should be made at the customer level, but data might not be available to estimate the required parameters on a per customer basis. Therefore, LTV calculations are often made for the "average" customer using average parameters. However, calculating the LTV of a group of customers assuming an average retention rate will technically not yield the correct average LTV. The reason is that the mean of a function of a variable "X" does not equal the function evaluated at the mean of X (i.e., $E[f(X)] \neq f(E[X])$). The correct way to calculate the average LTV of a group of customers is to determine each customer's parameters (e.g., retention rate), use them to calculate each customer's LTV, and then average. For this reason, even if the firm only needed to compute LTV at the segment level, it is preferable to compute individual LTV values and then average at the segment level.

Larger discount factors result in future profit being less "important" to the firm. Despite the importance of the discount factor in the calculation of LTV, there is very little systematic work on what value to use. In practice, one sees annual discount factors varying between 10% ($\delta = 0.10$) and 20% ($\delta = 0.20$), usually with little justification. We cover this issue in depth in Chapter 6.

$E(\tilde{V}_t)$ can be decomposed into revenues and costs. Specifically, $\tilde{V}_t = \tilde{R}_t - C_t$ where $\tilde{R}_t$ is the revenue generated by the customer in period t and C_t includes costs of goods, marketing and servicing. Little has been written about how to compute relevant costs for LTV models. For example, one very important issue is how to treat "fixed" versus "variable" costs. This topic will be covered in Chapter 6. We assume future costs are known, but revenues are random, so to compute *expected* lifetime value, we need to compute $E(\tilde{R}_t)$, expected revenue. To do this, we multiply the probability the customer is retained through period t (the survival rate in Table 5.1a) times the expected revenue generated during the period, given the customer has survived. Formally, $E(\tilde{R}_t) = P(\mathit{Survive\ until\ period\ t}) \cdot E(\tilde{R}_t|\mathit{Survive\ until\ period\ t}) = S_t \cdot E(\tilde{D}_t)$ where

S_t is the probability the customer survives until period t, and $\tilde{D}_t$ is a random variable equal to the revenue the customer generates during period t, given the customer survives until then.

Hazard models can be used to estimate S_t, and regression models can be used to estimate $E(\tilde{D}_t)$. A significant challenge is to incorporate control variables such as pricing and marketing contacts. We discuss these issues in Chapters 6 and 28.

5.3 The Two Primary LTV Models: Simple Retention and Migration

There are two primary models used to calculate LTV – simple retention and migration (Dwyer 1989; Berger and Nasr 1998). Simple retention models assume once the customer has attrited, the customer is lost to the company. Table 5.1a assumes a simple retention model. Migration models acknowledge that customers might migrate in and out of being a customer during the normal course of their lifetimes. Simple retention models are more applicable for industries such as financial services, B2B businesses, magazine subscriptions, and pharmaceutical drugs. Migration models are more applicable for industries such as retailing, catalogs, and consumer packaged goods.

5.3.1 Simple Retention Models

5.3.1.1 Calculating the Retention Rate by Direct Observation

As Table 5.1a illustrates, one of the most important parameters for the simple retention model is the retention rate, the probability the customer remains with the company, given the customer has not yet left the company. The simplest way to calculate a retention rate is by direct observation. Using its customer base for year 1, the firm can determine what percentage of these customers remained with the company in year 2. The resulting retention rate is often assumed to apply for all periods. The computation can be made more detailed by segmenting customers based on how long they have been customers or by various other demographic or behavioral variables.

The method just described is perhaps the most common way that retention rates are calculated in the real world. The problem is that they assume retention rates from the past will hold up in the future. They also provide little flexibility to calculate the probability customer will still be with the firm 11/2 years from now, or to calculate customer-level retention rates. Overcoming these deficiencies requires a model for which hazard models provide an ideal approach (see Chapter 15 for more detailed discussion of hazard models).

Table 5.3 LTV using a constant hazard rate

Example 1: Parameters
Hazard rate = 0.1
Revenue = $200
Cost = $100
Discount rate = 10%

LTV calculation:

Year	Hazard	Retention rate	Survival rate	Discount multiplier	Discounted expected profit
1	0	1	1.00	1.00	$100.00
2	0.1	0.9	0.90	0.91	$81.82
3	0.1	0.9	0.81	0.83	$66.94
4	0.1	0.9	0.73	0.75	$54.77
5	0.1	0.9	0.66	0.68	$44.81
6	0.1	0.9	0.59	0.62	$36.66
7	0.1	0.9	0.53	0.56	$30.00
8	0.1	0.9	0.48	0.51	$24.54
9	0.1	0.9	0.43	0.47	$20.08
10	0.1	0.9	0.39	0.42	$16.43

LTV after 10 years = $476.06
LTV over infinite horizon using Equation 5.2 = $550.00

Example 2: Parameters
Hazard rate = 0.2
Revenue = $200
Cost = $100
Discount rate = 10%

LTV calculation:

Year	Hazard	Retention rate	Survival rate	Discount multiplier	Discounted expected profit
1	0	1	1.00	1.00	$100.00
2	0.2	0.8	0.80	0.91	$72.73
3	0.2	0.8	0.64	0.83	$52.89
4	0.2	0.8	0.51	0.75	$38.47
5	0.2	0.8	0.41	0.68	$27.98
6	0.2	0.8	0.33	0.62	$20.35
7	0.2	0.8	0.26	0.56	$14.80
8	0.2	0.8	0.21	0.51	$10.76
9	0.2	0.8	0.17	0.47	$7.83
10	0.2	0.8	0.13	0.42	$5.69

LTV after 10 years = $351.49
LTV over infinite horizon using Equation 5.2 = $366.67

5.3.1.2 Using Hazard Models to Calculate LTV for a Simple Retention Model

Hazard models are used to compute S_t, the probability that a customer is still alive at (survived to) time t. Let $\tilde{T}$ be a random variable representing the time the customer attrites (dies) with probability density function f(t). The probability of attrition is $\mathrm{P}(\tilde{T} < t) = \mathrm{F}(t)$, where F($t$) is the cumulative distribution function: $\mathrm{F}(t) = \int_0^t f(x)dx$. The probability that a customer survives past time t is $S_t = \mathrm{P}(\tilde{T} \geq t) = 1 - \mathrm{F}(t) = \int_t^\infty f(x)dx$.

The hazard function is also very useful. It is the probability the customer attrites during the instantaneous period Δt *given the customer has remained with the firm up to period t*. It is defined as $h(t) = \frac{f(t)}{S(t)}$. The hazard functions can also be represented as $h(t) = \frac{d}{dx} \log S(t)$. Thus for a given survival function, there is a one-to-one relationship with a corresponding hazard function.

To show how the survival function is used in LTV calculations, we will begin with a very simple distribution for survival rates, the exponential distribution, where $\mathrm{f}(t) = \lambda e^{-\lambda t}$. The survival function for the exponential distribution is $S_t = 1 - \mathrm{F}(t) = e^{-\lambda t}$ and the hazard function is $h(t) = \frac{f(t)}{S(t)} = \lambda$. This means that if the lifetime of the customer follows an exponential distribution, the hazard is constant each period no matter how long the customer survives.

To make matters simpler, we will use the discrete version of the exponential distribution, the geometric distribution with parameter h. The hazard for the geometric distribution in any discrete time period is h and is constant. Let $r = 1 - h$ which is the retention rate. The survival function for τ periods after the initial period is r^τ. The value of a customer up to period τ is $\sum_{t=1}^{\tau} r^{t-1}(R_t - C_t)/(1+\delta)^{t-1}$ where R_t is revenue, δ is the discount rate and C_t is cost in period t.

Table 5.3 shows the computation of expected profit for the geometric assuming a hazard rate h per period and a retention rate each period of $r = 1 - h$. The table also shows the survival rate. The expected profit in each period is the survival rate times the discounted profit per period, which we assume to be known and constant. This case, where we assume constant retention rate and profit contribution, can be calculated using a simple formula:

$$LTV = (R - C)\frac{1+\delta}{1+\delta-r} \tag{5.2}$$

where δ is the discount rate, r is the retention rate, and R and C are the assumed known revenues and costs per period.[2]

[2] Assume revenues (R) and costs (C) are constant over time, and the discount factor is δ. Then

$$\begin{aligned} \mathrm{LTV} &= \sum_{t=1}^{\infty} \frac{(R-C)r^{t-1}}{(1-d)^{t-1}} = (R-C) + \frac{r(R-C)}{(1+\delta)} + \frac{r^2(R-C)}{(1+\delta)^2} \cdots \\ &= (R-C) \times (1 + d + d^2 \ldots) \end{aligned}$$

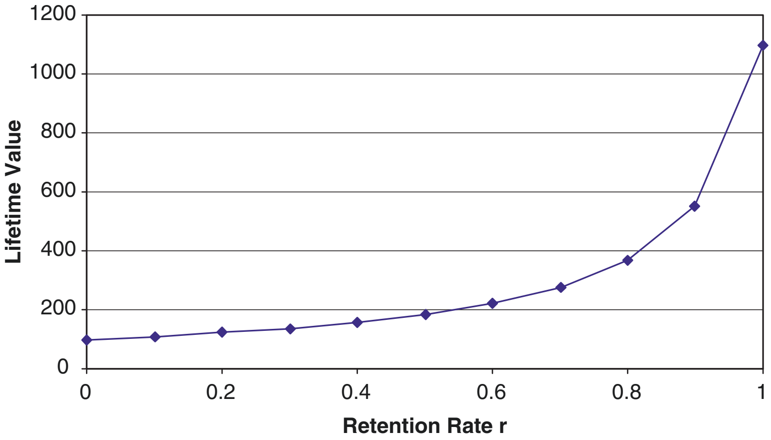

Fig. 5.1 Lifetime value as a function of retention rate – simple retention model.

Figure 5.1 shows the relationship between LTV and the retention rate using Equation 5.2. The relationship is convex and increases significantly as retention rate approaches one. This is the reason many authors argue small increases in the retention rates have a significant impact on LTV. What they do not include is the cost of changing the retention rate. It may be very costly to increase it from 0.90 to 0.95.

Other statistical distributions can be used to generate survival and hazard functions. For example, if one believes that customers have declining hazards, meaning that the longer the customers are with the firm, the lower the probability of attrition, then a Weibull distribution with specific parameters can be used. The Weibull distribution has a probability distribution function (p.d.f) of:

$$f(t) = \lambda\gamma(\lambda t)^{\gamma-1}e^{-(\lambda t)^{\gamma}} \tag{5.3}$$

The survival and hazard functions for the Weibull are respectively:

$$S(t) = e^{-(\lambda t)^{\gamma}} \tag{5.4a}$$

$$h(t) = \lambda\gamma(\lambda t)^{\gamma-1} \tag{5.4b}$$

The shape of the survival and hazard functions are determined by γ. If $\gamma < 1$, then the hazard function is decreasing over time, and if $\gamma > 1$, then it is increasing. If $\gamma = 1$, the Weibull become the exponential distribution with constant hazard λ. Figure 5.2 plots the hazard function for $\gamma = 0.7$ and $\gamma = 1.3$ (with $\lambda = 0.3$). The shape of the hazard is extremely useful for database marketers because it tells the decision maker whether the risk of customer attrition increases or decreases over time.

where $d = \frac{r}{(1+\delta)}$. Since $r < 1$ and $(1+\delta) > 1$, $d < 1$, we have an infinite geometric series. The sum of this series is $(R - C) \times \frac{1}{1-d} = (R - C) \times (\frac{1+\delta}{1+\delta-r})$.

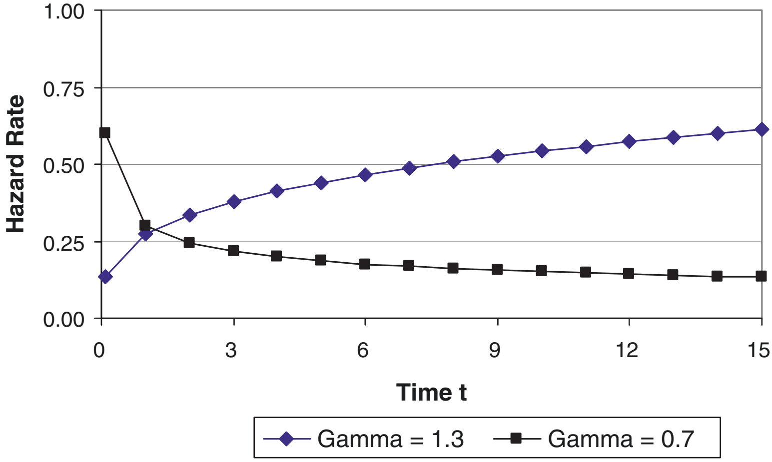

Fig. 5.2 Hazard function for Weibull with $\gamma = 0.7$ and 1.3 ($\lambda = 0.3$).

Researchers should try to explain the shape of the hazard function. For example, a decreasing hazard function may be due to heterogeneity across customers in their preference for the product and may not be due to changing hazard rates. Alternatively, as the customers uses the product or service, he or she increases preference or satisfaction or is locked in and that causes the decreasing hazard.

One can always transform continuous into discrete distributions.[3] Table 5.4 shows the results for a Weibull distribution with 10 years of data with $\gamma = 0.7$ and $\lambda = 0.3$. It shows a different pattern than one sees for a constant hazard rate model. For example in year 2 the retention rate increases to 0.81. In year 3 it is 0.84. The retention rate is the key statistic for computing LTV. By multiplying the retention rates over time, we compute the implied survival rate, i.e., the probability the customer is still active. Multiplying this times the \$100 sales per year yields expected sales per year.

Hazard models are a powerful tool for calculating lifetime value. They can be used to calculate customer-specific LTV because they can be extended to incorporate customer-level information such as demographics and marketing variables (Chapter 15). They are flexible and can allow a constant or time-varying retention rate. Hazard models can be estimated using commonly available statistical packages. See Chapter 15, Seetharaman and Chintagunta (2003), and Lawless (2003) for more details.

A key issue that affects the use of hazard models for LTV computations is that we may not know when a customer attrites. For many non-subscription businesses, the firm cannot determine if the customer attrites. For example,

[3] It is often useful to use discrete distributions because continuous time distributions frequently require numerical integration to compute LTV. It is easier to think of revenue streams in discrete intervals as well.

Table 5.4 Transforming a continuous hazard into a discrete hazard

Weibull parameters:
Gamma = 0.7
Lambda = 0.3
Sales per customer per year = $100

Time	Survival	True hazard rate	Estimated hazard rate	Estimated retention rate	Estimated survival rate	Expected sales per year
0.1	0.92	0.60	–	–	–	–
1	0.65	0.30	0.34	0.66	0.66	$65.88
2	0.50	0.24	0.27	0.73	0.48	$48.27
3	0.39	0.22	0.23	0.77	0.37	$37.24
4	0.32	0.20	0.21	0.79	0.30	$29.55
5	0.26	0.19	0.19	0.81	0.24	$23.89
6	0.22	0.18	0.18	0.82	0.20	$19.58
7	0.19	0.17	0.17	0.83	0.16	$16.22
8	0.16	0.16	0.16	0.84	0.14	$13.56
9	0.13	0.16	0.16	0.84	0.11	$11.41
10	0.12	0.15	0.15	0.85	0.10	$9.67
11	0.10	0.15	0.15	0.85	0.08	$8.23
12	0.09	0.14	0.14	0.86	0.07	$7.04
13	0.07	0.14	0.14	0.86	0.06	$6.05
14	0.07	0.14	0.14	0.86	0.05	$5.21

Survival rates and Hazard rates are computed from Equations 5.4a, b. The estimated hazard is computed by taking the change in survival rate and dividing it by the average survival rate for a given row and the row above it. This can then be compared to the actual hazard rate computed from the model.

The estimated retention rate is simply 1 – hazard rate which is computed from the estimated hazard rate. The estimated survival rate is the estimated retention rates multiplied up to the given point in time for which the survival rate is computed.

The results show that the estimated hazard is very close to the actual hazard rate.

a catalog company does not know when a customer has attrited. If the time of attrition is not known, hazard models cannot be estimated. Later we will discuss ways to incorporate the "death" process to estimate the probability of attriting (Sect. 5.4).

5.3.2 Migration Models

5.3.2.1 The Basic Customer Migration Model for Calculating LTV

The second common model for measuring LTV is the migration model, which as we will see in the next section, models LTV as a Markov Chain. The term migration model is used because the model allows customers to "migrate" among different states. The most common way of defining states is in terms of how recently the customer has bought from the company. This model acknowledges, in contrast to the simple retention model, that the customer might not purchase from the firm each period, but can skip one or more periods and still come back to purchase.

To operationalize this model, we define "recency state j" to mean that the customer last bought from the company j periods ago. We assign customers to recency states at the conclusion of each period. So if at the end of period 15 the customer is in recency state 2, that means the customer did not buy in period 15 but bought in period 14. Recency state 1 would mean that the customer bought in period 15. The key parameters that drive the migration model are the "recency probabilities"[4]:

p_j = Probability the customer purchases in the current period, given that the customer last purchased j periods ago, i.e., that the customer is classified in recency state j at the end of the previous period ($NR \geq j \geq 1$).[5]

The migration model reduces to the simple retention model if $p_j = 0$ for $j > 1$ because then if the customer is not retained, he or she cannot purchase again. We assume the recency probabilities do not change over time, i.e., we have no time subscript for p_j. This assumption could be relaxed at the cost of added complexity.

Table 5.5 illustrates the calculation of LTV using a migration model. We have four recency states ($NR = 4$), labeled 1, 2, 3, and ≥4. The state "≥4" signifies that it has been four or more periods since the customer has purchased. In Table 5.5, this customer has no chance of purchasing again – in Markov chain terminology, state ≥4 is an absorbing state. We have acquired the customer in period 1. Therefore, the customer is classified in recency state 1 at the end of period 1. The probability the customer purchases in period 2 is $p_1 = 0.5$. With probability $1 - p_1 = 0.5$, the customer does not purchase and hence moves to recency state 2 at the end of period 2. To compute the probability the customer purchases in period 3, we calculate P(customer in state 1) × P(Purchase|state 1) + P(customer in state 2) × P(Purchase|state 2) = $0.5 \times 0.5 + 0.2 \times 0.5 = 0.35$. The general pattern is that the customer moves to recency state 1 if he or she purchases in that period, or slips one recency state if he or she does not. Note that for the absorbing state ≥4, we assume there is no chance the customer will purchase again so the customer stays in that state. So for period 5, the probability the customer is in state ≥4 is P(Customer is in state ≥4 in period 4) + P(Customer does not purchase|Customer is in state 3 in period 4) $= 0.360 + (1 - 0.1) \times 0.200 = 0.540$.

---

[4] Note we generally follow the general development of Berger and Nasr in this section. See also also Dwyer (1989) and Calciu and Salerno (2002).

[5] Technically there is no upper limit to how many periods ago the customer might have purchased, but for computational purposes, we typically use an upper limit "NR". NR = "≥5" means that any customer who has not purchased in the past 5 or more periods would be classified in recency state ≥5.

Table 5.5 Migration model calculation

	Recency state (j)							
$j =$	1	2	3	≥4		Delta = 0.1		
$p_j =$	0.5	0.2	0.1	0	**Purch$_t$**	**Profit contribution if purchase**	**Expected profit**	**Discounted expected profit**
Period	**1**	**2**	**3**	**≥4**				
1	1.000	0.000	0.000	0.000	1.000	$100	$100.00	$100.00
2	0.500	0.500	0.000	0.000	0.500	$100	$50.00	$45.45
3	0.350	0.250	0.400	0.000	0.350	$100	$35.00	$28.93
4	0.265	0.175	0.200	0.360	0.265	$100	$26.50	$19.91
5	0.188	0.133	0.140	0.540	0.188	$100	$18.75	$12.81
6	0.134	0.094	0.106	0.666	0.134	$100	$13.43	$8.34
7	0.096	0.067	0.075	0.761	0.096	$100	$9.65	$5.45
8	0.069	0.048	0.054	0.829	0.069	$100	$6.92	$3.55
9	0.050	0.035	0.039	0.877	0.050	$100	$4.96	$2.31
10	0.036	0.025	0.028	0.912	0.036	$100	$3.56	$1.51
11	0.026	0.018	0.020	0.937	0.026	$100	$2.55	$0.98
12	0.018	0.013	0.014	0.955	0.018	$100	$1.83	$0.64
13	0.013	0.009	0.010	0.968	0.013	$100	$1.31	$0.42
14	0.009	0.007	0.007	0.977	0.009	$100	$0.94	$0.27
15	0.007	0.005	0.005	0.983	0.007	$100	$0.68	$0.18
							LTV = Total = $230.74	
					If extend calculation 100 periods, LTV = $231.08			

p_j = Probability customer buys in the current period, given the customer is in recency state j at the end of the previous period, i.e., last purchased j periods ago.
Delta = Discount factor = δ.
Purch$_t$ = Probability the customer buys in period t.

Table 5.5 shows that the sum of discounted expected profit after 15 periods is $230.74. One can see because of the declining values in later periods that this is close to the ultimate long-term LTV but not exact. Carrying out the calculation for 100 periods yields an LTV of $231.08.

Figure 5.3 shows sensitivity analyses based on the example in Table 5.5. The figure shows a convex relationship between recency probabilities p_2 and p_3 and LTV. These relationships could help a firm evaluate whether it would be worthwhile to attempt to induce customers who had not bought in say three periods to purchase this period.

Libai et al. (2002) describe a customer migration model for a European retailer. The models they use place customers into segments. Segment membership is dynamic. The authors argue that one way to increase customer equity is to increase the probability that a customer will move to a more profitable segment. By identifying key differences between segments, the firm can adjust the marketing and customer service mix for each segment. The concepts described by Libai et al. have the potential to link marketing mix actions to segment migration. The difficulty is creating the linkages. The authors do not describe the exact modeling methods they use. This becomes a research opportunity for academics and practitioners who have large customer databases and can develop the relevant methodology.

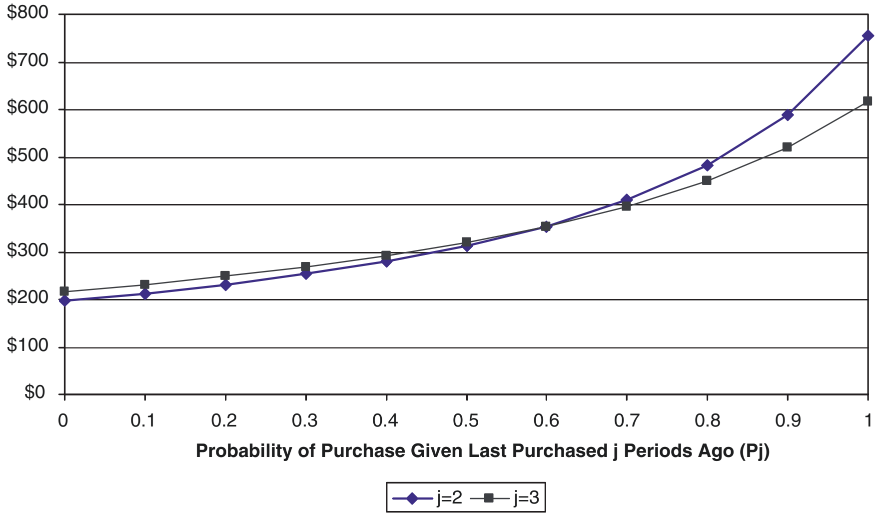

Fig. 5.3 Lifetime value as a function of recency purchase probabilities in a migration model.

5.3.2.2 Generalizing the Migration Model Using a Markov Chain Framework

Pfeifer and Carraway (2000) propose a Markov Chain framework that generalizes the calculations for the migration model. We use the same notation for p_j to signify the probability the customer purchases by the end of the current period, given the customer is in recency state j at the end of the previous period. We also label the recency states $\mathrm{R}_1, \mathrm{R}_2, \ldots, \mathrm{R}_{\mathrm{NR}}$. To simplify the exposition, we assume as in Pfeifer and Carraway that there are three possible (recency) states ($NR = 3$): bought last period (R_1), bought two periods ago (R_2), and bought three or more periods ago (R_3). The model of customer migration can be represented as a Markov chain with a 3×3 transition probability matrix, $\mathbf{P}$, as follows:

$$\begin{array}{c} \textbf{Period t+1} \\ \mathbf{P} = \textbf{Period t} \begin{array}{c} \\ R_1 \\ R_2 \\ R_3 \end{array} \begin{array}{c} \begin{array}{ccc} \mathrm{R}_1 & \mathrm{R}_2 & \mathrm{R}_3 \end{array} \\ \begin{bmatrix} p_1 & 1-p_1 & 0 \\ p_2 & 0 & 1-p_2 \\ p_3 & 0 & 1-p_3 \end{bmatrix} \end{array} \end{array} \tag{5.5}$$

Each element of $\mathbf{P}$ represents the probability the customer migrates from one state to another in a single period. Consider a customer in state R_1. The customer will or will not purchase in the current period with probabilities p_1 and $1 - p_1$ respectively. If the customer purchases, he or she remains in R_1. If not, the customer moves to R_2. The customer in state R_2 will or will not purchase at the current period with probabilities p_2 and $1 - p_2$. Finally, the customer in state R_3 will or will not purchase at the current period with probabilities p_3 and $1-p_3$. Some transitions have zero probability, such as R_2 to R_2. If a customer is in recency state R_2, he or she will either not purchase and go to recency state R_3 or purchase and go to R_1 but cannot stay in R_2.

Because of the property of the Markov Chain, we can easily calculate a t-step (t periods from now) transition matrix; that is, the matrix of probabilities of moving from one state to another after t periods. It is simply the matrix product of t one-step transition matrices, $\mathbf{P}^{\mathrm{t}}$. The $(i,\ j)$th element of the matrix $\mathbf{P}^t$ is the probability that the customer who begins at the ith state and will be at the jth state t periods later.

Assume that the firm earns MC if a customer purchases, and the firm spends M for a customer to repurchase. The profit vector $\mathbf{G}$ can be written as:

$$\mathbf{G} = \begin{bmatrix} MC - M \\ -M \\ -M \end{bmatrix} \tag{5.6}$$

The first row of $\mathbf{G}$ represents the profit contribution if the customer is in state R_1 (just purchased), the second row represents the profit contribution if the customer is in state R_2 (the customer did not purchase but the firm spent M on marketing to the customer), etc. Let $\Pi_{\mathbf{1}}$ = the profit after one period. Then,

$$\Pi_1 = \mathbf{PG} = \begin{bmatrix} p_1 MC - M \\ p_2 MC - M \\ p_3 MC - M \end{bmatrix} \tag{5.7}$$

The vector of the expected profit is $\mathbf{P}^2\mathbf{G}$ after two periods, and $\mathbf{P}^T\mathbf{G}$ after T periods. And the corresponding vector of the expected net present value that considers the discount factor d per period is $\mathbf{P}^1\mathbf{G}/(1+d)$ after one period, $\mathbf{P}^2\mathbf{G}/(1+d)^2$ after two periods, and $\mathbf{P}^T\mathbf{G}/(1+d)^T$ after T periods. Therefore, the vector of the total net present value from the period 0 to the period T is:

$$\mathbf{V}^T = \sum_{t=0}^{T} [(1+d)^{-1}\mathbf{P}]^t\mathbf{G} \tag{5.8}$$

Hence the vector of the total net present value for the infinite time horizon becomes

$$\mathbf{V}^\infty = \lim_{T\to\infty} \mathbf{V}^T = [\mathbf{I} - (1+d)^{-1}\mathbf{P}]^{-1}\mathbf{G} \tag{5.9}$$

where $\mathbf{I}$ is the identity matrix. The vector in Equation 5.9 is of particular interest because each element represents the expected lifetime value of a customer starting in state R_1, R_2, etc. The first element of $\mathbf{V}^\infty$ is of particular interest – it is the long-term value of a customer who starts out in state

R_1. This is the LTV of a just-acquired customer, because a just-acquired customer starts off in state R_1 (just purchased). The second element of $\mathbf{V}^{\infty}$ is the net present value of a customer who we currently observe to be in state R_2 (purchased two periods ago). In this way, we see that the Markov framework is a generalization of the migration model presented in the previous section.

To illustrate the connection between the "brute force" calculation in Table 5.5 and the matrix calculation, note the recency probabilities in Table 5.5 imply the following:

$$\mathbf{P} = \begin{bmatrix} .5 & .5 & 0 & 0 \\ .2 & 0 & .8 & 0 \\ .1 & 0 & 0 & .9 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5.10}$$

The matrix P is 4×4 because we have four states. The payoff matrix is simple:

$$\mathbf{G} = \begin{bmatrix} \$100 \\ \$0 \\ \$0 \\ \$0 \end{bmatrix} \tag{5.11}$$

The customer contributes \$100 if he or she purchases; else contributes \$0, and we are not considering any period-by-period marketing costs. To complete the model, note that the discount rate in Table 5.5 is 10%. From Equation 5.9, we therefore have:

$$\begin{aligned} \mathbf{V}^{\infty} &= \left\{ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \left(\frac{1}{1+0.1} \right) \begin{bmatrix} .5 & .5 & 0 & 0 \\ .2 & 0 & .8 & 0 \\ .1 & 0 & 0 & .9 \\ 0 & 0 & 0 & 1 \end{bmatrix} \right\}^{-1} \begin{bmatrix} \$100 \\ \$0 \\ \$0 \\ \$0 \end{bmatrix} \\ &= \begin{bmatrix} \$231.08 \\ \$57.29 \\ \$21.01 \\ \$0 \end{bmatrix} \end{aligned} \tag{5.12}$$

The first element of $\mathbf{V}^{\infty}$ is the LTV of the customer. So LTV of the customer is \$231.08, which is the number we obtain in Table 5.5 by extending the table to 100 periods ($\approx\infty$).

As another example, assume a customer generates \$40 marginal contribution (MC) whenever a purchase is made, and the firm mails to the customer unless the customer lapse to state ≥ 4, in which case the firm knows the customer will not purchase no matter what, so doesn't bother to mail a catalog. The mailing costs are \$4. The firm's discount rate is $d = 0.2$. The payoff matrix, $\mathbf{G}$, is then:

$$\mathbf{G} = \begin{bmatrix} MC - M \\ -M \\ -M \\ 0 \end{bmatrix} = \begin{bmatrix} 36 \\ -4 \\ -4 \\ 0 \end{bmatrix}$$

Next suppose the transition matrix is defined as

$$\mathbf{P} = \begin{bmatrix} .3 & .7 & 0 & 0 \\ .2 & 0 & .8 & 0 \\ .05 & 0 & 0 & .95 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

This transition matrix implies that if the customer is state R_1, he or she has a 0.3 chance of purchasing. If the customer is in state R_2, there is a 0.2 chance of a purchase but if the customer is in state R_4, the customer is the "absorbing" state "$\geq$4" and has no chance of purchasing.

Suppose we want to study the behavior of a new customer (in state R_1) to determine how likely the customer is to be in state R_1 after four purchase occasions? We can multiply the transition matrix three times and see what the purchase pattern will be. For the period after the first purchase, we see that:

$$\mathbf{P}^2 = \begin{bmatrix} .23 & .21 & .56 & 0 \\ .10 & .14 & 0 & .76 \\ .015 & .035 & 0 & .95 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and after the 4th period,

$$\mathbf{P}^4 = \begin{bmatrix} .0823 & .0973 & .1288 & .6916 \\ .037 & .0406 & .056 & .8664 \\ .0070 & .0080 & .0084 & .9766 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The matrix shows that a customer who began in state R_1 has a 0.0823 probability of being in R_1 (having just made a purchase) 4 periods later.

Another question that can be answered using the method just described is that we can examine the expected present value for each initial state after 4 periods, which is:

$$\prod_4 = \sum_{i=0}^{4} \frac{P^i G}{(1+d)^i} = \begin{bmatrix} \$49.40 \\ \$2.69 \\ (\$1.98) \\ 0 \end{bmatrix} \quad \text{where } P^0 G = G$$

The interpretation of the 3rd row element of Π_4 is that a customer beginning in recency state 3 has a negative expected present value after 4 periods. The firm should not mail individuals in this cell because the value is negative at time 0.

Pfeifer and Carraway show that the states do not have to be defined in terms of recency. They can be defined in terms of recency plus frequency, for example. In this case, R_1 might represent a customer who just bought and has bought once over the last year. R_2 might represent a customer who just bought and has bought twice over the last year. If we have four recency states, each with four frequency states, the transition matrix will be a 16×16, but the same machinery as described in Equations 5.6–5.9 would be applicable. Particularly interesting would be to experiment with different marketing policies, which would change the values of the **G** vector. Pfeifer and Carraway show how this might be done in the context of a catalog manufacturer.

In summary Pfeifer and Carraway's formulation of migration models as a Markov chain is a valuable generalization. It provides a framework for extending the definition of customer states, and for experimenting with various marketing policies. Their contribution is very practical and hence of value to managers.

5.4 LTV Models that Include Unobserved Customer Attrition

A series of LTV models have been developed that incorporate unobserved attrition (Schmittlein et al. 1987; Fader et al. 2004, 2005). They derive results such as the expected future number of remaining purchases or the customer's expected remaining lifetime with the firm, given the customer's past purchase history. These models therefore can be used to value customers in terms of the future number of purchases, lifetime duration, or lifetime value.

A fundamental notion in these models is the concept of whether the customer is "alive" or "dead." Customer attrition (or churn) is an important concern for many companies (Chapter 24). In contractual settings, such as subscriptions in the telecom, magazine, or cable industries, it is easy to determine when customers have attrited – they do not renew their contract. However, in many industries the customer has no written contract with the company so attrition is unobserved. For example, catalog companies are known to keep sending catalogs to customers who have not purchased in several years. Perhaps those customers have attrited (churned) and sending them catalogs is a fruitless investment. This issue is of prime importance to non-contractual businesses such as travel services, restaurants, retail, health care providers, and catalogers.

Determining whether a customer is alive or dead would at first seem trivial – if the customer has not bought in a long time, he or she has attrited. However, what if the customer has an erratic, infrequent buying pattern? Perhaps there is a hiatus in the customer's purchase pattern (e.g., when a customer is on vacation or has changed jobs) and again will buy from the firm without any remedial action.

The models developed to date focus on four key phenomena:

- The number of purchases in a given time period.
- Heterogeneity in the parameters of the purchase rate model.

- Customer lifetime, i.e., how long the customer is alive.
- Heterogeneity in the parameters governing the lifetime model.

Table 5.6 shows how these phenomena are modeled. There are three main models developed to date. Schmittleim et al. (1987) (SCM) and Fader et al. (2005) (FHL) model purchases as a Poisson process, whereas Fader et al. (2004) (FHB) model it as a Bernouli process. SCM model customer lifetime as an exponential distribution, whereas FHL and FHB model it as a geometric process. We will concentrate our discussion on SCM and FHL.

SCM's propose the following:

- *Purchase rate*: While alive, each customer purchases according to a Poisson process with parameter λ.
- *Heterogeneity in purchase rate process*: The parameter λ is distributed across customers according to a gamma distribution with parameters r and α, so that the mean λ is r/α and the variance is r/α^2.
- *Lifetimes*: Each customer's lifetime follows an exponential distribution with parameter μ. μ is the "death rate," i.e., the mean of the lifetime distribution is $1/\mu$.
- *Heterogeneity in lifetimes*: The parameter μ is gamma distributed across customers with parameters s and β, so that the mean μ is s/β and the variance is s/β^2.

The most debatable assumptions are the Poisson purchase rates and exponential lifetimes. Both assumptions entail the memoryless property. For example the number of purchases by an alive *individual* customer with known parameters purchasing in the next t units of time is independent of how many purchases he or she made in any previous period of time. The implied exponential distribution between purchases (given the customer is alive) means that the customer is most likely to make another purchase directly after the previous purchase. This might hold in certain industries but not those where purchasing builds up customer inventory and the customer does not purchase again until the inventory is depleted. In addition, both the time between purchases and customer lifetimes have the property that the mean equals the variance, which is not intuitive and appears to be restrictive. The exponential lifetime assumption implies that the modal customer behavior is to leave the firm fairly soon after being acquired. This might be a good assumption in some non-contractual settings.

SCM derive two important metrics: (1) the probability the customer is alive and (2) the expected time a customer will purchase in a period of length T^*. SCM derive that for a customer who has made x purchases over T time periods, with the last purchase being at time t, the probability the customer is still alive can be expressed as[6]:

[6] This assumes $\alpha > \beta$. The authors derive other formulas for other cases.

Table 5.6 Comparison of stochastic models of unobserved customer attrition (From Schmittlein et al. 1987; Fader et al. 2004, 2005)

	Method		
	Pareto/NBD (P/NBD)	**Beta geometric/NBD (BG/NBD)**	**Beta geometric/beta binomial (BG/BB)**
	Schmittlein et al. (1987)	Fader et al. (2005)	Fader et al. (2004)
Phenomena			
Customer alive	τ = lifetime $\sim$ Exp(μ)	p = probability "die" after each purchase	q = probability "die" after each purchase
Heterogeneity	$\mu \sim$ gamma(s, β)	$p \sim$ Beta(a, b)	$q \sim$ Beta(γ, δ)
Purchase rate	x = number of purchases in time $t \sim$ Poisson (λ)	x = number of purchases in time $t \sim$ Poisson (λ)	p = probability of purchase in each period
Heterogeneity	$\lambda \sim$ gamma(r, α)	$\lambda \sim$ gamma(r, α)	$p \sim$ Beta(α, β)
Summary			
Liftetimes	Exponential	Geometric	Geometric
Purchases	Poisson	Poisson	Bernoulli
Interpurchase	Exponential	Exponential	Geometric

$$P(Customer\ Alive) = \left\{1 + \frac{s}{s+x+s}\left[\left(\frac{\alpha+T}{\alpha+t}\right)^{r+x}\left(\frac{\beta+T}{\alpha+t}\right)^{s} \times F - \left(\frac{\beta+T}{\alpha+T}\right)^{s} F\right]\right\}^{-1} \quad (5.13)$$

where F is the Gaussian hypergeometric function[7] (Schmittlein et al. 1987, p. 6) with four parameters related to the four parameters that govern the model: $a_1 = r+x+s$, $b_1 = s+1$, $c_1 = r+x+s+1$ and $z_1(T) = \frac{\alpha-b}{\alpha+y}$. The expected quantity purchased in a period of length T^* is:

$$E[X^*|T^*] = \frac{(r+x)(\beta+T)}{(\alpha+T)(s-1)}\left[1 - \left(\frac{\beta+T}{\beta+2T}\right)^{s-1}\right] \times P(Customer\ Alive) \quad (5.14)$$

Another important calculation is the expected remaining time that the customer stays with the company, calculated at time t. SCM show that after taking into account heterogeneity, the remaining lifetime for a customer (τ) follows a Pareto distribution:

$$f(\tau|s,\beta) = \frac{s}{\beta}\left[\frac{\beta}{(\beta+\tau)}\right]^{s+1} \quad (5.15)$$

where s and β are the parameters of the gamma distribution of heterogeneity in death rate μ. The expected value of the Pareto distribution is

$$E[\tau|s,\beta] = \frac{\beta}{(s-1)} \quad (5.16)$$

The authors point out that since the death rate and purchase rate are assumed independent, purchases made up to time T have no impact on the remaining time we expect the customer to live. However, if the customer is active at time T, we do need to update the β parameter to $\beta+T$. Therefore, given the customer is still active at time T, the remaining lifetime follows a Pareto distribution with parameters $\beta+T$ and s. So the expected remaining lifetime for a customer with purchase history $\{x,t,T\}$ is:

$$E[remaining\ lifetime|x,t,T,\alpha,t,\beta,s] = \frac{(\beta+T)}{(s-1)}P(customer\ alive) \quad (5.17)$$

where $P(customer\ alive)$ is calculated using Equation 5.13.

The above discussion centers on the "Paredo/NBD" model developed by Schmittlein et al. More recently, this model has been extended by

[7] See Fader et al. (2005) for a method to approximate the Gaussian hypergeometric function.

Fader et al. (2005) referred to as FHL. This extension retains the original ideas of the Paredo/NBD model but is much easier to estimate. In fact, the authors provide Excel spreadsheets for estimating the models.

FHL's model is called the Beta Geometric/NBD, or BG/NBD model. It models lifetime as a geometric distribution rather than an exponential distribution. The customer has a probability p of becoming inactive after any transaction, so:

$$P(\textit{become inactive after } j^{th} \textit{ transaction}) = p(1-p)^{j-1} \tag{5.18}$$

The parameter p is analogous to the death rate μ in the Pareto/NBD model, and also is modeled to be heterogeneous across customers, following a beta distribution. Fader et al. (2004) derive formulas for $P(\textit{Alive})$ and $E(\textit{Number of purchases})$ analogous to those for the Pareto/NBD model.

FHL assume a consumer follows a Poisson process for purchasing. Hence the interpurchase time is exponential. Specifically, let λ equal the purchase rate and t_j be the time the customer purchases for the jth time. Then,

$$f(t_j|t_{j-1}, \lambda) = \lambda e^{-\lambda(t_j - t_{j-1})} t_j > t_{j-1} \geq 0 \tag{5.19}$$

As stated earlier, the probability an individual becomes inactive after the jth transaction is $p(1-p)^{j-1}$ where p is the probability the customer becomes inactive immediately after the jth purchase. These two equations then generate expressions of interest in computing LTV. The first is the *expected number of purchases* in a period of length t is:

$$E(X(t)|\lambda, p) = \frac{(1 - e^{-\lambda pt})}{p} \tag{5.20}$$

and the *probability the customer is alive at time* τ is:

$$P(\tau > t) = e^{-\lambda pt} \tag{5.21}$$

The critical parameters are λ and p. The higher p is, the fewer purchases and the less likely the customer is to be alive after t. This seems reasonable since the probability of dying should determine the number of purchases. In Equation 5.20 the limit as t goes to infinity is $1/p$ which is the expected number of purchases the customer will make over the long run. The probability the customer is alive decreases in p, which is also intuitively reasonable.

A potential problem is that as λ increases, the expected number of purchases increases because the numerator of Equation 5.20 increases. However, the expected time the customer remains alive declines as λ increases Equation 5.21. If a customer has a higher purchase rate, then he or she will have a lower probability of being alive at time τ. This follows from the assumption that the customer has probability p of dying after each purchase. FHL's model implicitly assumes the more times the customer purchases, the higher the probability the customer attrites. This is a questionable assumption.

Once FHL develop the individual customer model, they then study how heterogeneity in the purchase rate and probability of dying change the results. They use Gamma heterogeneity for λ and Beta heterogeneity for p.

$$f(\lambda|r,\alpha) = \frac{\alpha^r \lambda^{r-1} e^{-\lambda\alpha}}{\Gamma(\alpha)} \text{ for } \lambda > 0 \tag{5.22}$$

$\Gamma(\alpha)$ is the gamma function evaluated at α.

$$f(p|a,b) = \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)} 0 \leq p < 1 \tag{5.23}$$

where $\mathrm{B}(a,b)$ is the beta function which is equal to $\Gamma(a)\Gamma(b)/\Gamma(a+b)$.

The key expectations derived by FHL are:

$$E(X(t)|r,\alpha,a,b) = \frac{a+b-1}{a-1}\left[1-\left(\frac{\alpha}{\alpha+t}\right)^r {}_2F_1(r,b;a+b-1;\left(\frac{t}{\alpha+t}\right)\right] \tag{5.24}$$

where ${}_2F_1(\bullet)$ is the Gaussian hypergeometric function. This is the expected number of purchases for the whole customer base over time. Obviously, the hypergeometric function makes it more difficult to understand the intuition behind the results.

The other expectation of interest is the expected number of transactions for an individual with a specific observed behavior characterized by the number of purchases made, x (frequency), the last time a purchase was made t_x (recency) and the length of the interval, T.

Let $E(Y(t)|X = x, t_x, T, r, \alpha, a, b)$ equal the expected number of transactions for a time period of length t given the number of prior purchases, the last observed purchase t_x and the end of the interval T. Then,

$$E(Y(t)|X = x, t_x, T, r, \alpha, a, b) =$$
$$\frac{\frac{\mathrm{a+b+x-1}}{a-1}\left[1-\left(\frac{\alpha+T}{\alpha+T+t}\right)^{r+x} {}_2F_1\left[\left(r+x, b+x; a+b+x-1; \left(\frac{t}{\alpha+T+t}\right)\right]\right.\right.}{1+\delta_{>0}\frac{a}{b+x-1}\left(\frac{\mathrm{a+T}}{\mathrm{a+t_x}}\right)^{\mathrm{r+x}}} \tag{5.25}$$

While this expression appears to be complex, FHL argue that the Gausian hypergeometric function can be approximated using Excel. FHL test their model versus the Pareto/NBD and find it to be equivalent.

For a fixed future time interval, FHL's model will provide an estimate of LTV. They compute Equation 5.25 for a customer who makes x purchases with his or her last purchase at time t_x, with the end of the base period (period before computing future LTV) being T and the future period of length t, what is the expected number of purchases. This times the expected margin gives the undiscounted future LTV. To transform FHL (or SMC) into an LTV calculation with discounting requires setting the model up with discrete time intervals and creating conditional probabilities of staying alive (the hazard function) for their models.

To see how FHL's model behaves as a function of its parameters, we will provide a relatively simple version of their model by looking at one consumer and avoiding the complexity added by studying heterogeneity. The three key quantities that FHL (and SCM) provide and which are extremely useful in LTV modeling are:

$E[X(t)|\lambda, p]$ = the number of transactions in a period of length t,
$P(\tau > t|\lambda, p)$ = the probability the customer will be alive after period t, and
$E[Y(t)|x, t_x, T, \lambda, p]$ = the expected number of transactions in the period t to T, for an individual with observed behavior $X = x$, t_x, where t_x is the time of the last purchase in the interval $[0, t]$.

From above, λ is the purchase rate per period and p is the probability the customer attrites after a purchase. For the example, we will use several values of the parameters and time periods to show how the above quantities change. We will use the following formulas to compute the quantities described above:

$$\mathrm{E[X(t)|\lambda, p]} = \frac{1 - e^{-\lambda pt}}{p} \tag{5.26a}$$

$$P(\tau > t|\lambda, p) = e^{-\lambda pt} \tag{5.26b}$$

$$\begin{aligned} E[Y(t)|X &= x, t_x, T, \lambda, p) \\ &= \frac{p^{-1}(1-p)^x \lambda^x e^{-\lambda T} - p^{-1}(1-p)^x \lambda^x e^{-\lambda(T+pt)}}{L(\lambda, p|X = x, t_x, T)} \end{aligned} \tag{5.26c}$$

with

$$L(\lambda, p|X = x, t_x, T) = (1-p)^x \lambda^x e^{-\lambda T} + \delta_{x>0}\, p(1-p)^{x-1} \lambda^x e^{-\lambda t_x} \tag{5.26d}$$

The above expressions are easily computed in Excel. We will start with the expected number of purchases in given time interval t. We will use two values of λ and two values of p to contrast the expected number of purchases. Table 5.7**a** provides the values. It shows, as should be obvious, as p (the death probability) increases, the number of purchases decreases and as λ increases; so does the expected number of purchases. For a simple Poisson model with $\lambda = 0.1$, which represents .1 purchase per month and a time interval of 24 months, the expected number of purchases would be $\lambda t = 2.4$. For FHL, we see that the expected number of purchases is 2.13 purchases for $t = 24$ (24 months). As p increases, we see that the expected number of purchases decreases. For $p = 0.25$ and $\lambda = 0.1$, the expected value decreases to 1.8. Thus, the larger the value of p, the more the model diverges from a Poisson purchase rate model.

An interesting result is shown for the probability that a customer is alive which is displayed in Table 5.7**b**. It shows that as λ, the purchase rate,

Table 5.7 Calculating the expected number of purchases and the probability the customer is alive (From Fader et al. 2005)

	Death rate(p)	
	0.10	**0.25**
(a) Expected number of purchases over 24 periods ($t = 24$)		
Purchase rate (λ)		
0.10	2.13	1.80
0.25	4.51	3.10
(b) Probability customer is alive after 24 periods		
Purchase rate (λ)		
0.10	0.787	0.549
0.25	0.549	0.223
(c) Expected number of purchases over 24 periods, given past purchase history ($t = 24, t_x = 5, T = 10, x = 3$)		
Purchase rate (λ)		
0.10	1.80	1.16
0.25	3.25	1.43

(d) Expected number of purchases over 24 periods, varying recency $(t_x)(t = 24, T = 10, x = 3, p = 0.1)$

	Purchase rate (λ)	
	0.10	**0.30**
Recency (t_x)		
3	1.74	2.69
5	1.80	3.42
7	1.85	4.03
9	1.90	4.46

increases, the probability of being alive decreases for fixed p. As discussed earlier, this is a counter-intuitive result.

Another result worth noting is the expected number of purchases in a given time interval, 24 months, *given past purchase behavior*. This is shown in Table 5.7**c**. This quantity is different than the expected number of purchases because it is conditional on the time of the last purchase. The table shows that the longer since the last purchase, the customer is expected to make fewer purchases. Intuitively, if a customer has a high purchase rate but does not purchase in a fixed interval, the likelihood increases that the customer has died. Hence, for a fixed number of purchases, x, the less recent t_x, the time of the last purchase is, the smaller the number of purchases. This makes intuitive sense and is consistent with the findings from recency modeling.

Table 5.7**d** shows that as λ increases, the impact of t_x is much greater. The change in the expected number of purchases has much greater differences from low to high values of t_x. Again this result makes intuitive sense.

In conclusion, the computations from FHL's model show that many of their results are reasonable except for the assumption that each time a customer makes a purchase, the probability of attriting is constant. This leads to counter-intuitive results regarding the probability of attriting as a function of the purchase rate. It may also lead to poor fits if their model is applied at the individual level rather than at the aggregate level after allowing for heterogeneity.

The next extension, due to Fader et al. (2004) (FHB) builds on the BG/NBD modeling the process entirely in discrete time. The key change is moving from a Poisson to a Bernoulli purchase process. In each period, the customer has the probability, p, of purchasing. Periods are independent and p is constant, analogous to the stationarity and independence assumptions of the Poisson. In fact, a Bernoulli process becomes a Poisson process as we move from discrete to continuous time. The parameter p is heterogeneous across customers according to a Beta distribution. The death process for the customer follows the same geometric distribution as FHL, but the unit is the time period, not the transaction. So each time period, the customer has a probability q of dying, and q is heterogeneous across customers according to a Beta distribution. The authors call this model the beta-geometric/beta-binomial (BG/BB) model. Again, they derive formulas for $P(Alive)$ and *E[Number of Purchases]* as in the other cases.

In summary, the stochastic models developed by SCM, FHL, and FHB provide models of lifetime value when customer attrition is unobserved. These models can be used to calculate customer lifetime value estimates both on average (analogous to a simple retention model) and for individual customers with purchase histories (particularly the number of purchases and time of the last purchase). However, one needs to be careful when computing it for individuals because the amount of data per individual may be very small. The models are rich yet remarkably simple – Table 5.6 states all that is needed to define each model – and with the extensions of Fader et al. relatively easy to estimate.

There are two main areas for building on these models. The first is to allow for inter-relationships between the purchase and death processes, and especially between these processes and quantity purchased. Surely the parameters governing these processes should be correlated across customers. One would assume for example that customers with high purchase rates would have lower death rates.

The second extension is to incorporate marketing variables in these models. The parameters of the model could be made functions of marketing variables. This would require for example a hierarchical Bayesian framework, for example:

$$p \sim Beta(\alpha, \beta) \tag{5.27a}$$

$$\alpha \sim Normal(\mu, \sigma^2) \tag{5.27b}$$

$$\mu = f(\textit{marketing effort}) \tag{5.27c}$$

The real complexity would come in predicting individual-level response in terms of purchase rates or *P[Alive]* as a function of marketing efforts. This endeavor, however, would be well worth it because it would turn what now is a *ceteris paribus* lifetime value model into a *potential* lifetime value model.

5.5 Estimating Revenues

The next major requirement in computing LTV is estimating revenues per customer per period. There are many possible ways to make these estimates.

5.5.1 Constant Revenue per Period Model

One computes the average revenue per customer for all periods and then uses this measure for revenue per customer per period. This is very easy to use but very naïve and likely to be unrealistic. It is often the case that revenue increases over time.

5.5.2 Trend Models

One can calculate the trend in revenue per period per customer from the initial customer acquisition period to the end of the customer's purchase series. We might use segments or aggregate across customers. The trend model then is used to capture the pattern of customer revenues over time. The trend can be modeled using a constant growth rate or a growth curve which asymptotes to a specific value or other shapes depending upon the revenue data pattern.

5.5.3 Causal Models

Revenue can be estimated using causal models in which the dependent variable is the log of spending (to avoid negative predictions) and the independent variables are causal variables such as price and other relevant variables that should predict spending. The decision maker or researcher can use historical values or patterns in causal variables to serve as the independent variables in the predictions. The problem with causal models is that while they fit may the data, they must predict future spending. To overcome this problem we can create scenarios to understand how different values of independent variables affect spending levels. Then the firm can decide upon which scenario(s) best fit likely firm behavior.

5.5.4 Stochastic Models of Purchase Rates and Volume

One could use the distribution of purchase volume across consumers to predict purchase volume for individual customers. The prediction will be a weighted average of the customer's historical purchase volume and the mean for all customers, the weights being determined by how many observations are available for the customer. (See Columbo and Jiang (1999) and the description of the model in Chapter 12.)