PETER S. FADER, BRUCE G.S. HARDIE, and KA LOK LEE*

The authors present a new model that links the well-known RFM (recency, frequency, and monetary value) paradigm with customer lifetime value (CLV). Although previous researchers have made a conceptual link, none has presented a formal model with a well-grounded behavioral "story." Key to this analysis is the notion of "iso-value" curves, which enable the grouping of individual customers who have different purchasing histories but similar future valuations. Iso-value curves make it easy to visualize the interactions and trade-offs among the RFM measures and CLV. The stochastic model is based on the Pareto/NBD framework to capture the flow of transactions over time and a gamma-gamma submodel for spend per transaction. The authors conduct several holdout tests to demonstrate the validity of the model's underlying components and then use it to estimate the total CLV for a cohort of new customers of the online music site CDNOW. Finally, the authors discuss broader issues and opportunities in the application of this model in actual practice.

RFM and CLV: Using Iso-Value Curves for Customer Base Analysis

The move toward a customer-centric approach to marketing, coupled with the increasing availability of customer transaction data, has led to an interest in both the notion and the calculation of customer lifetime value (CLV). At a purely conceptual level, the calculation of CLV is a straightforward proposition: It is simply the present value of the future cash flows associated with a customer (Pfeifer, Haskins, and Conroy 2005). Because of the challenges associated with forecasting future revenue streams, most empirical research on lifetime value has actually computed customer profitability solely on the basis of customers' prior behavior, but to be true to the notion of CLV, measures should look to the future not the past. A significant barrier has been the ability to model future revenues appropriately, particularly in the case of a "noncontractual" setting (i.e., where the time at which customers become "inactive" is unobserved) (Bell et al. 2002; Mulhern 1999).

*Peter S. Fader is Frances and Pei-Yuan Chia Professor of Marketing, Wharton School, University of Pennsylvania (e-mail: faderp@wharton.upenn.edu). Bruce G.S. Hardie is Associate Professor of Marketing, London Business School (e-mail: bhardie@london.edu). Ka Lok Lee is Market Research Analyst, Catalina Health Resource (e-mail: kaloklee@alumni.upenn.edu). The authors thank Albert Bemmaor, Paul Farris, Phil Pfeifer, the three anonymous *JMR* reviewers, and seminar participants at the Yale School of Management and University of Auckland for their helpful comments. The second author acknowledges the support of the London Business School Centre for Marketing.

Many researchers and consultants have developed "scoring models" (i.e., regression-type models) that attempt to predict customers' future behavior (see, e.g., Baesens et al. 2002; Berry and Linoff 2004; Bolton 1998; Malthouse 2003; Malthouse and Blattberg 2005; Parr Rud 2001). In examining this work, we note that measures of customers' prior behavior are key predictors of their future behavior. In the direct marketing literature, it is common practice to summarize customers' prior behavior in terms of their recency (time of most recent purchase), frequency (number of prior purchases), and monetary value (average purchase amount per transaction), that is, their RFM characteristics.

However, there are several problems with these models, especially when attempting to develop CLV estimates:

- Scoring models attempt to predict behavior in the next period, but when computing CLV, we are interested not only in Period 2 but also in Periods 3, 4, 5, and so on. It is not clear how a regression-type model can be used to forecast the dynamics of buyer behavior well into the future and then tie it all back into a "present value" for each customer.
- Two periods of purchasing behavior are required: Period 1 to define the RFM variables and Period 2 to arrive at values of the dependent variable(s). It would be better if it were possible to create predictions of future purchasing behavior using only Period 1 data. More generally, it would be nice to be able to leverage all of the available data for model calibration purposes without using any of it to create a dependent variable for a regression-type analysis.

- The developers of these models ignore the problem that the observed RFM variables are only imperfect indicators of underlying behavioral traits (Morrison and Silva-Risso 1995).[1] They fail to recognize that different "slices" of the data will yield different values of the RFM variables and, therefore, different scoring model parameters. This has important implications when the observed data from one period are used to make predictions of future behavior.

The way to overcome these general problems is to develop a formal model of buyer behavior that is rooted in well-established stochastic models of buyer behavior. In developing a model based on the premise that observed behavior is a realization of latent traits, we can use Bayes' theorem to estimate a person's latent traits as a function of observed behavior and then predict future behavior as a function of these latent traits. At the heart of our model is the Pareto/NBD framework (Schmittlein, Morrison, and Colombo 1987) for the flow of transactions over time in a noncontractual setting. An important characteristic of our model is that RFM measures are sufficient statistics for an individual customer's purchasing history. That is, rather than including RFM variables in a scoring model simply because of their predictive performance as explanatory variables, we formally link the observed measures to the latent traits and show that no other information about customer behavior is required to implement our model.

Before conveying our analytical results, we offer a brief exploratory analysis to set the stage for our model development. Consider the purchasing of the cohort of 23,570 people who made their first purchase at CDNOW in the first quarter of 1997. We have data on their initial and subsequent (i.e., repeat) purchases through the end of June 1998 (for further details about this data set, see Fader and Hardie 2001). For presentational simplicity, we initially examine only the relationship between future purchasing and recency/frequency; we introduce monetary value subsequently. We first split the 78-week data set into two periods of equal length and group customers on the basis of frequency (where x denotes the number of repeat purchases in the first 39-week period) and recency (where t_x denotes the time of the last of these purchases). We then compute the average total spend for each of these groups in the following 39-week period.[2] These data appear in Figure 1.

It is clear that recency and frequency each has a direct and positive association with future purchasing, and there may be some additional synergies when both measures are high (i.e., in the back corner of Figure 1). However, despite the large number of customers we used to generate this graph, it is still sparse and, therefore, somewhat untrustworthy. Several "valleys" in Figure 1 are simply due to the absence of any observations for particular combinations of recency and frequency. To "fill in the blanks," we must abstract away from the observed data and develop a formal model.

An alternative view of the same relationship appears in Figure 2. This contour plot, or "30,000-foot view" of Figure 1, is an example of an "iso-value" plot; each curve in the graph links customers with equivalent future value despite differences in their prior behavior. Again, it is easy to understand how greater recency and greater frequency are correlated with greater future purchasing, but the jagged

Figure 1
AVERAGE TOTAL SPEND IN WEEKS 40–78 BY RECENCY AND FREQUENCY IN WEEKS 1–39

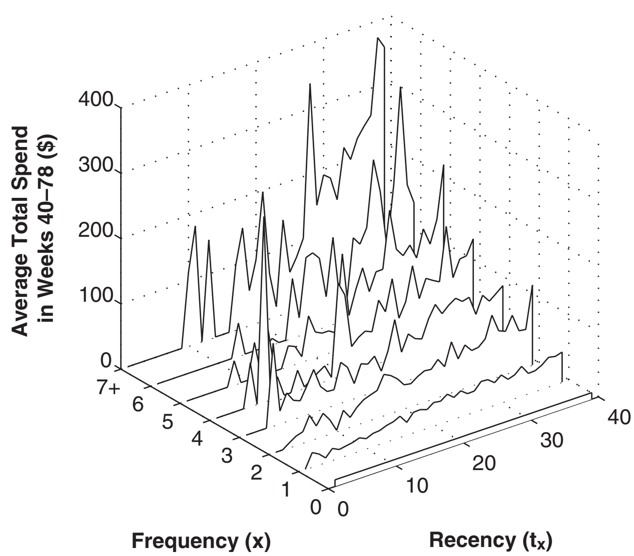
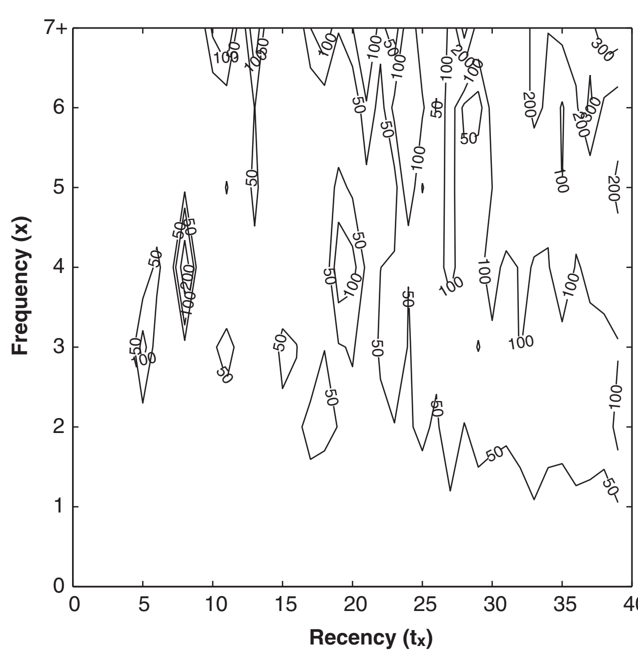


Figure 2
CONTOUR PLOT OF AVERAGE WEEK 40–78 TOTAL SPEND BY RECENCY AND FREQUENCY



[1]This same problem occurs with Dwyer's customer migration model (Berger and Nasr 1998; Dwyer 1989) and its extensions (Pfeifer and Carraway 2000).

[2]We have removed the purchasing data for ten buyers who purchased more than $4,000 worth of CDs across the 78-week period. According to our contacts at CDNOW, these people are probably unauthorized resellers who should not be analyzed in conjunction with ordinary customers. Therefore, this initial exploratory analysis is based on the purchasing of 23,560 customers.

nature of these curves emphasizes the dangers of relying solely on observed data in the absence of a model.

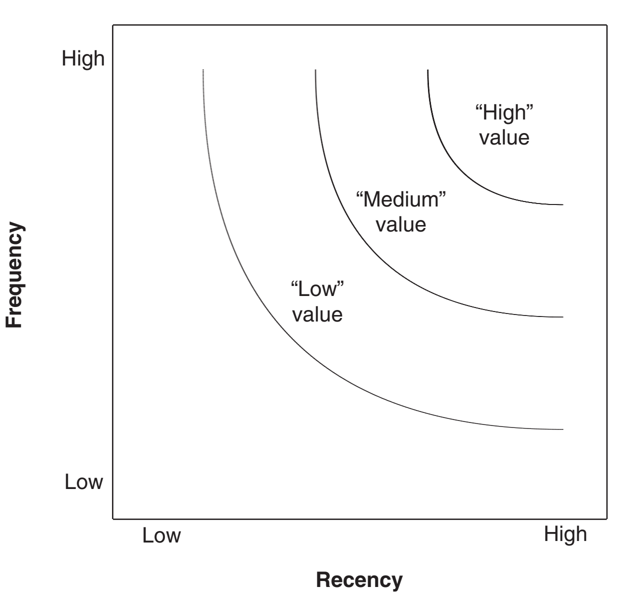
We should also note that in using such plots to understand the nature of the relationship between future purchasing and recency/frequency, the patterns observed will depend on the amount of time for which the customers are observed (i.e., the length of Periods 1 and 2).

Despite these limitations of using the kind of data summaries in Figures 1 and 2, there is still clear diagnostic value in the general concept of iso-value curves for customer base analysis. We can extract the main patterns observed in Figure 2 into the stylized version that appears in Figure 3. The shape of these curves should seem fairly intuitive, and they are reasonably consistent with the coarse summary in Figure 2. A formal model will enable us to understand these relationships better, including possible exceptions to the simple curves shown here. Furthermore, it will enable us to create plots that do not depend on the amount of time for which the customers are observed. (A parallel is Schmittlein, Cooper, and Morrison's [1993] work on "80/20 rules" in purchasing, in which they use a model to create a time-invariant measure of customer concentration.)

Overall, the creation and analysis of iso-value curves is an excellent way to summarize and evaluate the CLV for an entire customer base. It can help guide managerial decision making and provide accurate quantitative benchmarks to gauge the "return on investment" for programs that companies use to develop and manage their portfolios of customers.

In the next section, we describe the model that underpins our effort to link RFM and CLV, one for which RFM characteristics are sufficient statistics for an individual customer's purchasing history. After briefly discussing the model results and some holdout validations to assess its efficacy, we discuss the iso-value curves. Then, we use our new model to compute CLV for the entire cohort of CDNOW customers. We conclude with a brief summary of our work and a discussion of promising future research directions.

Figure 3
HYPOTHESIZED RELATIONSHIP BETWEEN RECENCY/FREQUENCY AND CLV



LINKING RFM WITH FUTURE PURCHASING

The challenge we face is how to generate forward-looking forecasts of CLV. At the heart of any such effort is a model of customer purchasing that accurately characterizes buyer behavior and therefore can be trusted as the basis for any CLV estimates. Ideally, such a model would generate these estimates using only simple summary statistics (e.g., RFM) without requiring more detailed information about each customer's purchasing history.

In developing our model, we assume that monetary value is independent of the underlying transaction process. Although this may seem counterintuitive (e.g., frequent buyers might be expected to spend less per transaction than infrequent buyers), our analysis lends support for the independence assumption. This suggests that 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 can then rescale this number of discounted expected transactions (DET) by a monetary value "multiplier" to yield an overall estimate of lifetime value:

$$\text{CLV} = \text{margin} \times \text{revenue/transaction} \times \text{DET}. \tag{1}$$

We first develop our submodel for DET. Then, we introduce a separate submodel for expected revenue per transaction. Although we assume independence for these two processes, there are still some notable relationships (particularly between frequency and monetary value) that we explore subsequently.

The Pareto/NBD model that Schmittlein, Morrison, and Colombo (1987) developed has proved to be a popular and powerful model in explaining the flow of transactions in a noncontractual setting. Reinartz and Kumar (2000, 2003) provide excellent illustrations. This model is based on the following general assumptions about the repeat buying process:

- Customers go through two stages in their "lifetime" with a specific firm: They are active for some period of time, and then they become permanently inactive.
- While customers are active, they can place orders whenever they want. The number of orders a customer places in any given time period (e.g., week, month) appears to vary randomly around his or her underlying average rate.
- Customers (while active) vary in their underlying average purchase rates.
- The point at which a customer becomes inactive is unobserved by the firm. The only indicator of this change in status is an unexpectedly long time since the customer's transaction, and even this is an imperfect indicator; that is, a long hiatus does not necessarily indicate that the customer has become inactive. There is no way for an outside observer to know for sure (thus the need for a model to make a "best guess" about this process).
- Customers become inactive for any number of reasons; thus, the unobserved time at which the customer becomes inactive appears to have a random component.
- The inclination for customers to "drop out" of their relationship with the firm is heterogeneous. In other words, some cus-

tomers are expected to become inactive much sooner than others, and some may remain active for many years, well beyond the length of any conceivable data set.
- Purchase rates (while a customer is active) and drop out rates vary independently across customers.

Translating these general assumptions into specific mathematical assumptions results in the Pareto/NBD model.

The only customer-level information that this model requires is recency and frequency. The notation used to represent this information is $(X = x, t_x, T)$, where x is the number of transactions observed in the time interval (0, T] and t_x $(0 < t_x \leq T)$ is the time of the last transaction. In other words, recency and frequency are sufficient statistics for an individual customer's purchasing history.

Schmittlein, Morrison, and Colombo (1987) derive expressions for several managerially relevant quantities, including

- E[X(t)], the expected number of transactions in a time period of length t, which is central to computing the expected transaction volume for the whole customer base over time;
- $P(\text{“active”}|X = x, t_x, T)$, the probability that a customer with observed behavior $(X = x, t_x, T)$ is still active at time T; and
- $E(Y(t)|X = x, t_x, T)$, the expected number of transactions in the future period (T, T + t] for a customer with observed behavior $(X = x, t_x, T)$.

How can we use this model to compute DET (and therefore CLV)? Drawing on standard representations, we could compute the number of DET for a customer with observed behavior $(X = x, t_x, T)$ as

$$\text{DET} = \sum_{t=1}^{n} \frac{E[Y(t)|X = x, t_x, T] - E[Y(t-1)|X = x, t_x, T]}{(1+d)^t},$$

where the numerator is the expected number of transactions in period t and d is the discount rate.

However, this expression is rather cumbersome, and the analyst is faced with two standard decisions that anyone performing CLV calculations faces (Blattberg, Getz, and Thomas 2001): (1) what time horizon to use in projecting sales and (2) what time periods to measure (e.g., year, quarter). Furthermore, this expression ignores the specific timing of the transactions (i.e., early versus late in each period), which could have a significant impact on DET.

We could determine the time horizon by generating a point estimate of when the customer becomes inactive by finding the time at which $P(\text{“active”}|X = x, t_x, T)$ crosses below some predetermined threshold (cf. Reinartz and Kumar 2000, 2003; Schmittlein and Peterson 1994). However, the probability that the customer is active is already embedded in the calculation of $E(Y(t)|X = x, t_x, T)$, so this approach goes against the spirit of the model.

Given that these calculations are based on the Pareto/NBD model, a more logical approach would be to switch from a discrete-time formulation to a continuous-time formulation (as is often done in standard financial analyses) and compute DET (and thus CLV) over an infinite time horizon. Standing at time T, we compute the present value of the expected future transaction stream for a customer with purchase history $(X = x, t_x, T)$, with continuous compounding at rate of interest δ. In the Appendix, we derive the DET expression for the Pareto/NBD model:

$$\text{(2)} \quad \text{DET}(\delta|r, \alpha, s, \beta, X = x, t_x, T) = \frac{\alpha^r \beta^s \delta^{s-1} \Gamma(r + x + 1)\Psi[s, s; \delta(\beta + T)]}{\Gamma(r)(\alpha + T)^{r+x+1} L(r, \alpha, s, \beta|X = x, t_x, T)},$$

where r, α, s, and β are the Pareto/NBD parameters; $\Psi(\cdot)$ is the confluent hypergeometric function of the second kind; and $L(\cdot)$ is the Pareto/NBD likelihood function, which is given in Equation A1. The derivation of this expression for DET is a new analytical result and is central to our CLV estimation.

Adding Monetary Value

Until this point, "customer value" has been characterized in terms of the expected number of future transactions, or DET to be more precise. In the end, however, we are more interested in the total dollar value across these transactions (from which we can then calculate a customer's profitability). For this, we need a separate submodel for dollar expenditure per transaction.

We specify a general model of monetary value in the following manner:

- The dollar value of a customer's given transaction varies randomly around his or her average transaction value.
- Average transaction values vary across customers but do not vary over time for any given individual.
- The distribution of average transaction values across customers is independent of the transaction process.

In some situations (e.g., the purchasing of a product that a customer can hold in inventory), some analysts may expect to observe a relationship between transaction timing and quantity/value. However, for the situation at hand, we assume that recency/frequency and monetary value are independent. We explore the validity of this assumption subsequently.

Schmittlein and Peterson (1994) assume that the random purchasing around each person's mean is characterized by a normal distribution and that the average transaction values are distributed across the population according to a normal distribution. This implies that the overall distribution of transaction values can be characterized by a normal distribution.

Our initial empirical analysis is based on a 1/10th systematic sample of the whole cohort (2357 customers), using the first 39 weeks of data for model calibration and holding out the second 39 weeks of data for model validation. (After we complete these validations, we will use the entire 78 weeks for calibration to generate the iso-value curves.) Table 1 reports basic descriptive statistics on average repeat transaction value for the 946 people who made at least one

Table 1

SUMMARY OF AVERAGE REPEAT TRANSACTION VALUE PER CUSTOMER (WEEKS 1–39)

	$		$
Minimum	3	Mean	35
25th percentile	16	Standard deviation	30
Median	27	Mode	15
75th percentile	42	Skewness	4
Maximum	300	Kurtosis	17

repeat purchase in Weeks 1–39. The large differences in the mean, median, and mode indicate that the distribution of observed individual means is highly skewed to the right. This suggests that the unobserved heterogeneity in the individual-level means cannot be characterized by a normal distribution.

We could modify Schmittlein and Peterson's (1994) model to allow for this skewness by assuming that the underlying average transaction values follow a lognormal distribution across the population. If we are going to modify their model, we should also replace the individual-level normal distribution (that characterizes random purchasing around the person's mean) with one that has a nonnegative domain. An obvious choice would be the lognormal distribution. However, because there is no closed-form expression for the convolution of lognormals, we are unable to derive a model that has a closed-form expression for the distribution of a customer's average observed transaction value (from which we could then make inferences about the customer's true mean transaction rate, a quantity that is central to our CLV calculations). Therefore, we consider the gamma distribution, which has similar properties to those of the lognormal (albeit with a slightly thinner tail), adapting the gamma-gamma model that Colombo and Jiang (1999) originally proposed.

To help understand the role of the monetary value submodel, we first address a logical question: Given that we observe an average value across x transactions, denoted by m_x, why do we need a formal model for monetary value at all? The answer is that we cannot necessarily trust the observed value of m_x as our best guess of each customer's true underlying average transaction value E(M). For example, suppose that the mean expenditure across all customers across all transactions is \$35, but Customer A has made one repeat purchase totaling \$100. What value of E(M) should we use for our future projection for Customer A? Should we assume that $E(M) = m_1 = \$100$, or should we "debias" our estimate down toward the population mean? As we observe more repeat purchases for each customer, we expect that the observed value of m_x will become increasingly accurate as an estimate of the true mean, E(M).

For a given customer with x transactions, let $z_1, z_2, \ldots, z_x$ denote the dollar value of each transaction. We assume that the value of each transaction varies randomly around the customer's (unobserved) mean transaction value E(M). The customer's average observed transaction value $m_x = \Sigma_{i=1}^{x} z_i/x$ is an imperfect estimate of E(M). Our goal is to make inferences about E(M) given m_x, which we denote as $E(M|m_x, x)$. It is clear that $m_x \to E(M)$ as $x \to \infty$, but this could be a slow process.

We assume that the Z_i are i.i.d. gamma variables with shape parameter p and scale parameter ν. Using two standard relationships involving the gamma distribution, where

1. the sum of x i.i.d. gamma (p, ν) random variables is distributed gamma with shape parameter px and scale parameter ν, and
2. a gamma(px, ν) random variable multiplied by the scalar 1/x is itself distributed gamma with shape parameter px and scale parameter νx,

it follows that the individual-level distribution of m_x is given by

$$f(m_x|p, \nu, x) = \frac{(\nu x)^{px} m_x^{px-1} e^{-\nu x m_x}}{\Gamma(px)}.$$

To account for heterogeneity in the underlying mean transaction values across customers, we assume that the values of ν are distributed across the population according to a gamma distribution with shape parameter q and scale parameter γ. We assume that the parameter p is constant across customers, which is equivalent to assuming that the individual-level coefficient of variation is the same for all customers ($CV = 1/\sqrt{p}$). Taking the expectation of $f(m_x|p, \nu, x)$ over the distribution of ν leads to the following marginal distribution for m_x:

$$f(m_x|p, q, \gamma, x) = \frac{\Gamma(px+q)}{\Gamma(px)\Gamma(q)} \frac{\gamma^q m_x^{px-1} x^{px}}{(\gamma + m_x x)^{px+q}}. \tag{3}$$

To arrive at an expression for our desired quantity, $E(M|m_x, x)$, we use Bayes' theorem to derive the posterior distribution of ν for a customer with an average spend of m_x across x transactions:

$$g(\nu|p, q, \gamma, m_x, x) = \frac{(\gamma + m_x x)^{px+q} \nu^{px+q-1} e^{-\nu(\gamma + m_x x)}}{\Gamma(px+q)},$$

which is itself a gamma distribution with shape parameter px + q and scale parameter $\gamma + m_x x$. It follows that the expected average transaction value for a customer with an average spend of m_x across x transactions is

$$\begin{aligned} E(M|p, q, \gamma, m_x, x) &= \frac{(\gamma + m_x x)p}{px+q-1} \\ &= \left(\frac{q-1}{px+q-1}\right)\frac{\gamma p}{q-1} + \left(\frac{px}{px+q-1}\right)m_x. \end{aligned} \tag{4}$$

This is a weighted average of the population mean, $\gamma p/(q - 1)$, and the observed average transaction value, m_x. It should be clear that larger values of x will lead to less weight being placed on the population mean and more weight being placed on the observed customer-level average of m_x. We illustrate this "regression-to-the-mean" phenomenon in more detail when we subsequently revisit the monetary value submodel.

Assessing the Independence of the Monetary Value Assumption

The assumption that the distribution of average transaction values across customers is independent of the transaction process is central to the model of customer behavior we use to link RFM with CLV. Just how valid is this assumption?

Using the transaction data for the 946 people who made at least one repeat purchase in Weeks 1–39 (of a sample of 2357 customers), we find that the simple correlation between average transaction value and the number of transactions is .11. The magnitude of the correlation is largely driven by one outlier: a customer who made 21 transactions in the 39-week period, with an average transaction value of \$300. If we remove this observation, the correlation between average transaction value and the number of transactions drops to .06 ($p = .08$).

In Figure 4, we use a set of box-and-whisker plots to summarize the distribution of average transaction value, broken down by the number of repeat purchases in the first 39 weeks. Although we can recognize the slight positive correlation, the variation within each number-of-transactions group dominates the between-group variation.

Thus, although there is a slight positive correlation between average transaction value and the number of transactions, we do not believe that it represents a substantial violation of our independence assumption. In the next section, we provide some additional evidence that this modest relationship has no discernable impact on the performance of our model in a holdout setting.

MODEL VALIDATION

Before we use the preceding expressions to explore the relationship between RFM and CLV for the case of CDNOW, it is important to verify that the Pareto/NBD submodel for transactions and the gamma-gamma submodel for monetary value each provides accurate predictions of future buying behavior. A careful holdout validation such as this is missing from virtually all CLV-related literature, but it must be conducted before we can make any statements about the model's ability to estimate CLV. This analysis is based on the previously noted 1/10th systematic sample of the cohort (2357 customers), using the first 39 weeks of data for model calibration and holding out the second 39 weeks of data for model validation. (In the next section, however, we use all 78 weeks of data for our CLV analysis.)

We briefly summarize the key statistical results of the fit/forecasting performance of the Pareto/NBD. We refer the interested reader to a more detailed validation of the model in a separate analysis that Fader, Hardie, and Lee (2005b) conducted.

The maximum likelihood estimates of the model parameters are $\hat{r} = .55$, $\hat{\alpha} = 10.58$, $\hat{s} = .61$, and $\hat{\beta} = 11.67$. Using the expression for $E[X(t)]$, we compute the expected number of transactions for the whole group of 2357 customers for each of the 78 weeks (see Figure 5). We note that the Pareto/NBD model predictions accurately track the actual (cumulative) sales trajectory in both the 39-week calibration period and the 39-week forecast period, underforecasting by less than 2% at week 78.

Given our desire to use the Pareto/NBD model as a basis for the computation of CLV, we are more interested in our ability to predict individual-level buying behavior in the forecast period (Weeks 40–78) conditional on prior behavior (purchasing in Weeks 1–39). We compute this using the expression for $E(Y(t)|X = x, t_x, T)$ for each of the 2357 customers. In Figure 6, we report these numbers along with the average of the actual number of transactions in the forecast period, broken down by the number of repeat purchases in

Figure 4

THE DISTRIBUTION OF AVERAGE TRANSACTION VALUE BY NUMBER OF TRANSACTIONS

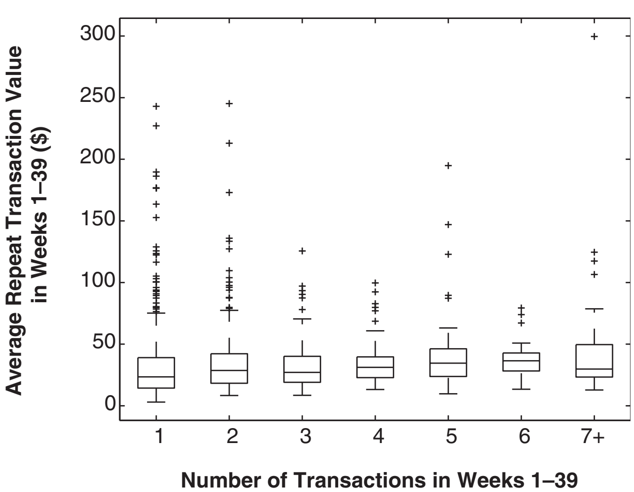


Figure 5

TRACKING CUMULATIVE REPEAT TRANSACTIONS

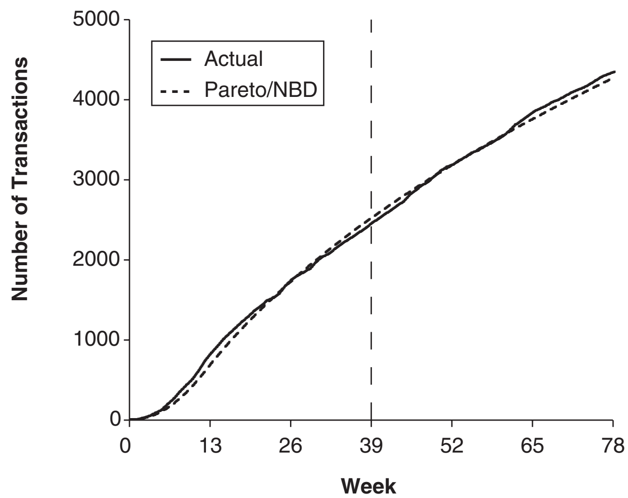
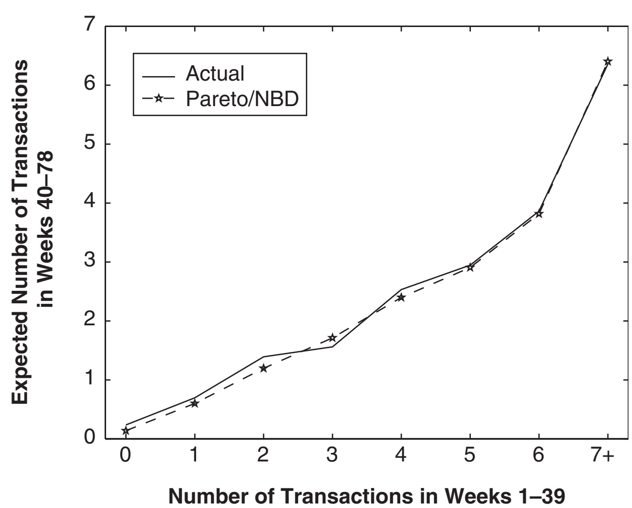


Figure 6

CONDITIONAL EXPECTATIONS OF PURCHASING (WEEKS 40–78)



the first 39 weeks. Note that for each frequency level (x), we average over customers of differing recency (i.e., different values of t_x). Our ability to closely track the sales data in a holdout period (at both the aggregate and the individual level) gives us confidence in the Pareto/NBD model as the basis for our upcoming CLV calculations.

The next step is to validate the monetary value model. The maximum likelihood estimates of the model parameters are $\hat{p} = 6.25$, $\hat{q} = 3.74$, and $\hat{\gamma} = 15.44$. The theoretical mean transaction value differs from the mean observed average repeat transaction value per customer by a mere $.09. To visualize the fit of the model, we compute the implied distribution of average transaction value across people using Equation 3 and compare it with a nonparametric density of the observed average transaction values in Figure 7.

The fit is reasonable. However, the theoretical mode of $19 is greater than the observed mode of $15, which corresponds to the typical price of a CD at the time the data were collected. The model is not designed to recognize the existence of threshold price points (e.g., prices ending in .99), so this mismatch is not surprising. The fit of the model could be improved by adding additional parameters, but for the purposes of this article, we choose to forgo such tweaks to maintain model parsimony.

A stronger test of our model for monetary value is to combine it with the Pareto/NBD model for transactions and examine the quality of the predictions of individual-level total spend in the forecast period (Weeks 40–78) of the data set. For each customer, we compute the expected average transaction value conditioned on his or her calibration period frequency and monetary value using Equation 4. We compute two sets of conditional expectations of forecast period monetary value: We obtain the first by multiplying each person's conditional expectation of per-transaction value by his or her actual number of forecast period transactions; we obtain the second by multiplying the person's conditional expectation of per-transaction value by his or her predicted number of forecast period transactions, conditional on the calibration period recency and frequency (i.e., the numbers summarized in Figure 6). These different projections enable us to test each subcomponent of the model separately and together.

In Figure 8, we report these two sets of conditional expectations along with the actual average total value of customers' forecast period transactions, broken down by calibration period frequency (i.e., number of repeat transactions). For each x, we average over customers of differing recency (values of t_x) and calibration period monetary value.

Our comparison of the "expected|actual x_2" numbers with the actual numbers can be viewed as a clean test of the monetary value submodel. In comparing these predictions with the actual numbers, we do not find any particular bias that would lead us to question the assumption of stationarity in the average buying rates or the assumption that average transaction value is independent of frequency. Our comparison of the expected|expected x_2 numbers with the actual numbers can be viewed as a test of the combined Pareto/NBD + gamma-gamma submodels. The combined submodels provide good conditional predictions of expected total monetary value in the forecast period; thus, we have confidence in using this model to make overall predictions of lifetime value outside the range of the observed data.

CREATING AND ANALYZING ISO-VALUE CURVES

Now that we have demonstrated the substantial validity of the Pareto/NBD submodel for transactions and the gamma-gamma submodel for monetary value, we can use all 78 weeks of data to make predictions about CLV. As we previously noted, our ability to use all the data is a significant advantage that helps distinguish a well-specified stochastic model from a more traditional scoring model; however, it is not immediately clear how the results of a standard two-period scoring model with RFM variables as key predictors can be projected beyond the observed data.

Figure 7

OBSERVED VERSUS THEORETICAL DISTRIBUTION OF AVERAGE TRANSACTION VALUE ACROSS CUSTOMERS

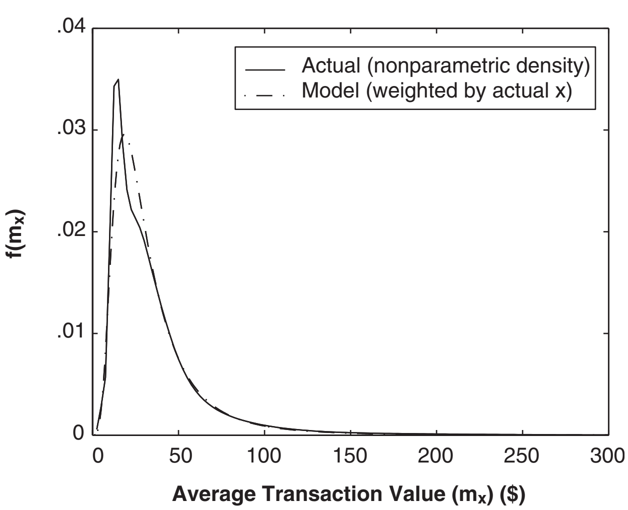
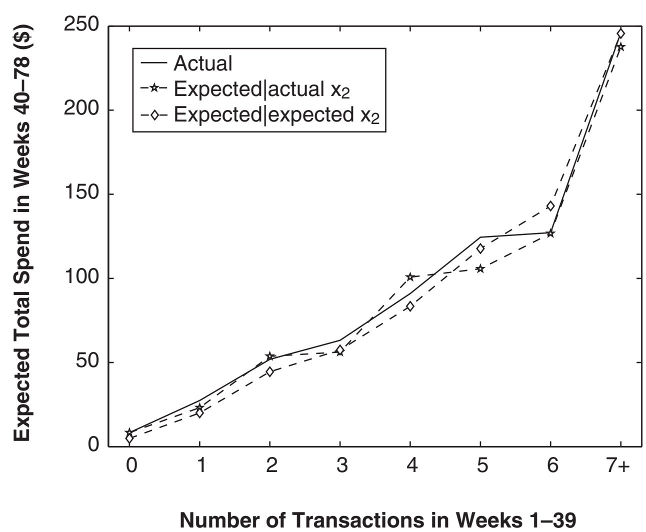


Figure 8

CONDITIONAL EXPECTATIONS OF MONETARY VALUE



We reestimate the models using all 78 weeks of data, and as a final validity check, we compare this "full" model with the 39-week version we used in the previous section. Fortunately, the model remains stable as we double the length of the calibration period. If we take the new 78-week Pareto/NBD parameters and plug them back into the 39-week log-likelihood (LL) function, we notice only a small decrease in model fit (from –9608 to –9595). For the gamma-gamma model, the degree of stability is remarkable: The 78-week parameters provide a 39-week LL of –4661, compared with the optimal 39-week LL of –4659. This provides strong support for our assumption that the submodel governing monetary value is stable over time.

We begin our iso-value analysis by focusing on the relationship between DET and recency/frequency. We then reintroduce monetary value to complete the picture. To obtain the iso-value curves for DET, we evaluate Equation 2 for all recency and frequency combinations ($t_x = 0, 1, \ldots, 78$; $x = 0, 1, \ldots, 14$).[3] The assumed annual discount rate is 15%, which implies a continuously compounded rate of $\delta = .0027$. The CLV estimates appear as a "waterfall" plot in Figure 9.

With the exception of $x = 0$,[4] DET is an increasing function of recency. However, note that there is a strong interaction with frequency. For low-frequency customers, there is an almost linear relationship between recency and DET. However, this relationship becomes highly nonlinear for high-frequency customers. In other words, for customers who have made a relatively large number of transactions in the past, recency plays a much greater role in the determination of CLV than for customers who have made infrequent purchases in the past.

The iso-value curves that appear in Figure 10 further shed light on this complex relationship. For the high-value customers (i.e., the upper-right area of Figure 10), the iso-value curves reflect the basic shape suggested by Figure 3, but the lower-value regions suggest that the iso-value lines begin to bend backwards. This may seem highly counterintuitive: Someone with frequency of $x = 7$ and recency of $t_x = 35$ has an approximate DET of 2, the same as someone with a lower (i.e., worse) frequency of $x = 1$ and recency of $t_x = 30$. In general, for people with low recency, higher frequency seems to be a bad thing.

To resolve this apparent paradox, consider the two hypothetical customers in Figure 11. If we were certain that both customers were still active in Week 78, we would expect Customer B to have a greater value of DET, given the higher number of prior purchases, but the pattern of purchases strongly suggests that Customer B is no longer

[3]For both computational and presentational simplicity, we perform these calculations with T = 77.86 (i.e., for a hypothetical customer who made his or her first purchase at CDNOW on January 1, 1997).

[4]For the case of x = 0 (i.e., no repeat purchases), recency has no meaning. In line with the work of Schmittlein, Morrison, and Colombo (1987), $t_x = 0$ for such cases.

Figure 9
DET AS A FUNCTION OF RECENCY AND FREQUENCY

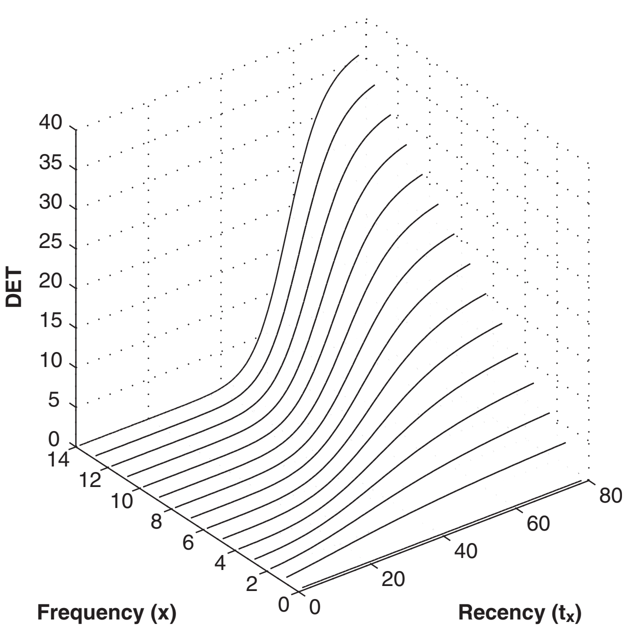


Figure 10
ISO-VALUE REPRESENTATION OF DET

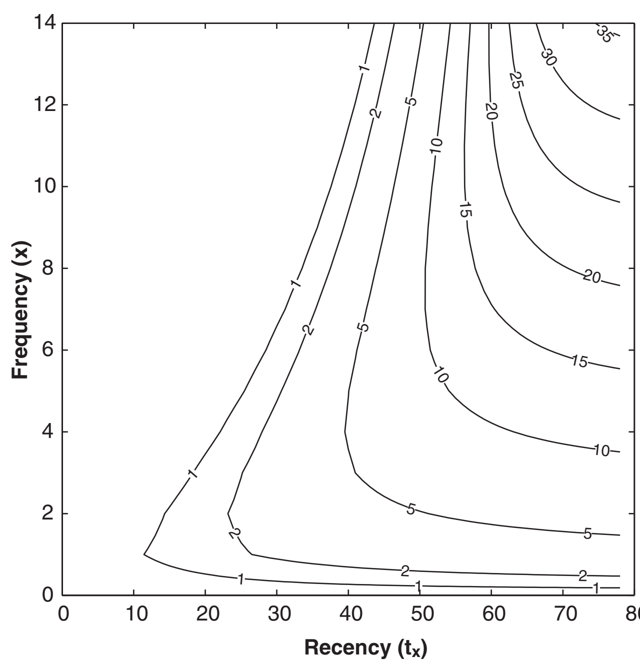
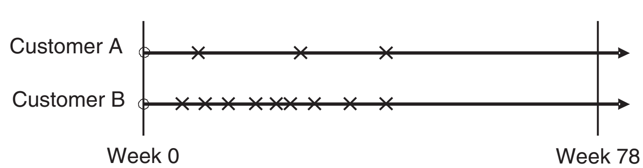


Figure 11
THE "INCREASING FREQUENCY" PARADOX



active. Conversely, Customer A has a lower underlying purchase rate, so it is more likely that he or she is still active in Week 78. The net effect is that the DET for Customer A is 4.6 and the DET for Customer B is 1.9.

Therefore, on further reflection, the existence of backward-bending iso-value curves makes sense. In our data set, we observe quite a few customers with purchasing histories that are qualitatively similar to those of Customer B, so it is important to capture this phenomenon.

These curves and the overall clarity provided by Figure 10 demonstrate the usefulness of using a formal model to understand the relationship between DET and recency/frequency. The sparseness we observe in the actual data (Figure 2), even though we used data for 23,560 customers, makes it difficult to visualize these vital relationships properly. Furthermore, the observed shape of the relationship does not fully reflect the true relationship (just as Schmittlein, Cooper, and Morrison's [1993] work on "80/20 rules" in purchasing demonstrates that observed concentration is not the true concentration).

Furthermore, the backward-bending iso-value curves emphasize the importance of using a model with sound behavioral assumptions rather than an ad hoc regression approach. The use of a regression-based specification, which is used in many scoring models, would likely miss this pattern and lead to faulty inferences for a large portion of the recency/frequency space. In contrast, it is reassuring that a straightforward four-parameter model such as the Pareto/NBD can capture such a rich and varying set of behavioral patterns.

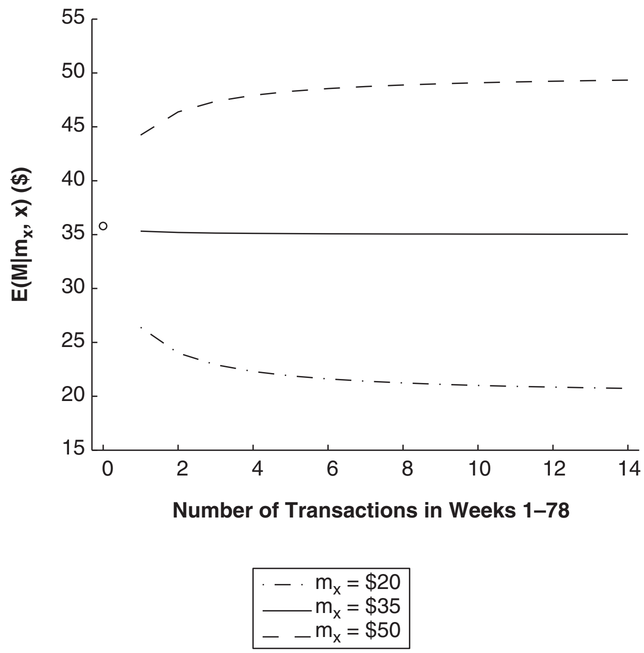
Having established some worthwhile relationships involving recency and frequency, it is now time to bring monetary value into the picture so that we can move from DET to CLV as our focal outcome variable. How do we augment our estimates of DET (e.g., Figure 9) with our predictions of monetary value to arrive at the estimate of CLV for all customers based on their purchase history (i.e., RFM) to date? Given our assumption of independence between the transaction and the monetary value processes, it is tempting simply to multiply the DET (as we predicted given recency and frequency) by the person's observed average monetary value (m_x), but this would ignore the "regression-to-the-mean" phenomenon we discussed previously and formalized in Equation 4. Instead, we need to take into account the number of transactions that customers made and use that information about their transaction history to derive a weighted average between their prior purchasing and the overall population tendencies. We illustrate this in Figure 12 for three different values of average observed transaction values (m_x): \$20, \$35, and \$50.

If there are no repeat purchases in Weeks 1–78, our best guess is that the person's average transaction value in the future is simply the mean of the overall transaction value distribution. This mean for the full 78-week period is \$36, which we identify by the circle in Figure 12. Therefore, our prediction of lifetime value for a customer with zero repeat purchases would be 36 times the DET value for x = 0 (times the gross margin), as we report in Figure 9.

With one observed repeat purchase, the best guess for E(M) moves toward the observed value of m_x (i.e., the amount of that single transaction) but not all the way. Not until the customer has made seven or eight transactions can

Figure 12

DET MULTIPLIERS AS A FUNCTION OF FREQUENCY AND AVERAGE TRANSACTION VALUE



we trust that the observed value of m_x serves as a reasonably accurate estimate of his or her true underlying average transaction value, E(M).

Finally, in translating from the dollar amount of each transaction to the financial value that the firm actually gains from each purchase, we assume a gross contribution margin of 30%. We have no information about the actual margins for CDNOW, but this number seems reasonably conservative and is consistent with the margins that other researchers (e.g., Reinartz and Kumar 2000, 2003) use. Choosing a different margin will change the values reflected in the subsequent iso-value curves, but it should not affect any of the main patterns observed in those figures.

We are now in a position to show the relationship between all three behavioral components (RFM) and CLV. From an analytical standpoint, we substitute Equation 2 and Equation 4 into Equation 1 along with our assumed margin of 30%. Given an assumed annual discount rate of 15%, the estimates of the four Pareto/NBD model parameters (r, α, s, β), and the three gamma-gamma model parameters (p, q, γ), the analyst can simply enter the observed values for RFM (t_x, x, and m_x, respectively) to derive an expected CLV for the customers who share those behavioral characteristics. In and of itself, the ability to obtain (presumably) accurate estimates for CLV from these basic inputs is a significant contribution from this work.

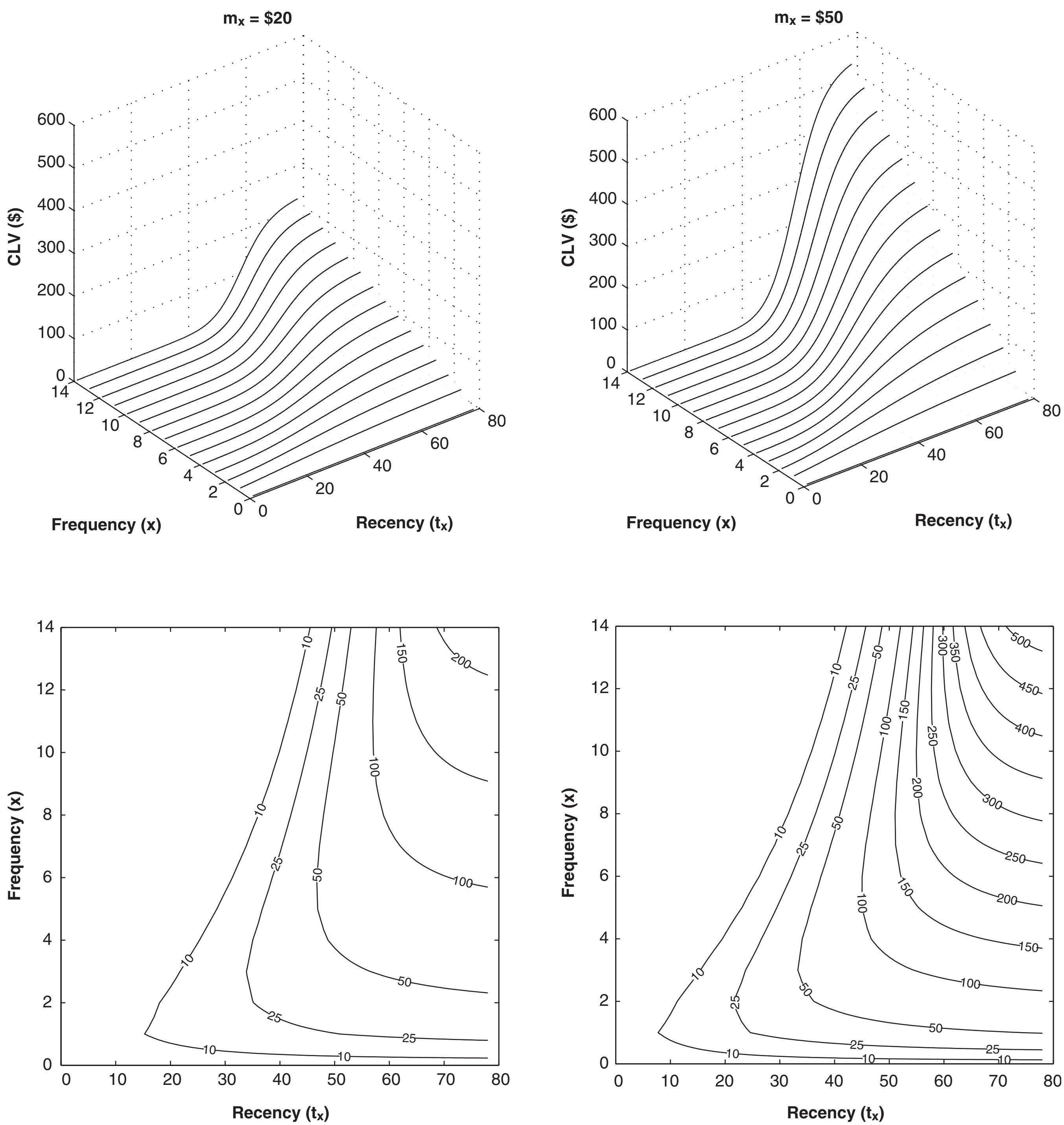
From a graphical standpoint, a complete view of this relationship would require us to move from the previous three-dimensional plots (Figures 1 and 9) to a four-dimensional representation. Although there are some high-dimensional visualization techniques that can provide such a summary, we did not find any to be satisfactory for the data at hand. Instead, we rely on the same type of recency/frequency plots we used previously, but we allow monetary value to vary at different levels across plots.[5]

In Figure 13, we show the waterfall and iso-value contour plots for two different levels of observed average transaction value (m_x): We use $20 and $50 as in Figure 12. At first

[5]We could bring monetary value in as a primary dimension. Indeed, we considered a variety of alternative plots (i.e., recency and monetary value by frequency, and frequency and monetary value by recency), but we found that there was a great deal of redundancy (and confusion) when we attempted to examine the data along different dimensions.

Figure 13

CLV AS A FUNCTION OF RECENCY AND FREQUENCY FOR AVERAGE TRANSACTION VALUES OF $20 AND $50



glance, these waterfall plots show some meaningful similarities to each other and to the earlier plots for DET only (Figures 9 and 10); recency and frequency each has a strong main effect on CLV, they interact positively when both are at high levels, and the flat area on the left-hand side of each graph reflects the "increasing frequency" paradox we discussed previously.

For the most part, these plots are essentially rescaled versions of each other (reflecting the assumption of independence between recency/frequency and monetary value), but the regression-to-the-mean patterns in Figure 12 are also a factor, particularly at lower levels of frequency. As x increases in each plot, the increases in CLV are altered by the diverging curves in that figure. For the case of m_x = \$20, the lower curve in Figure 12 shows that the changes associated with increasing frequency are less than linear, but for the case of m_x = \$50, the upper curve in Figure 12 shows that there is a greater-than-linear multiplier effect. These differences can be observed when comparing the lower-right portions of the two iso-value graphs.

Another way to understand and appreciate the usefulness of our method for computing CLV is to combine the model-driven RFM–CLV relationship with the actual RFM patterns observed in our data set. In Figure 14, we show the distribution in the recency/frequency space for the 11,506 customers (of 23,560) who made at least one repeat purchase at CDNOW in the 78-week observation period. Note that we have reversed the directions of the recency and frequency axes (for this figure only) to make it possible to observe the patterns here. This figure shows that the majority of repeating customers had only one repeat transaction; from the mass of customers at the back of the distribution, it is clear that many of these single repeat purchases occurred early in the observation period.

Essentially, our goal is to integrate this customer distribution along with the iso-value curves to derive a sense of overall CLV for the customer base. To enhance the clarity and interpretability of this combination, we group customers on the basis of their RFM characteristics. This enables us to "close the loop" with traditional RFM segmentation analyses to show how our model can be used for target marketing purposes.

We set aside the 12,054 customers who made no repeat purchases over the 78-week observation period. We assigned each of the remaining 11,506 customers an RFM code in the following manner: We sorted the list of customers in descending order on recency, and we coded the customers in the top tercile (most recent) R = 3, in the second tercile R = 2, and in the third tercile (least recent) R = 1. We then sorted the entire list in descending order on frequency; we coded members of the top tercile (highest number of transactions) F = 3, and so forth. Finally, we sorted the customer list in descending order on average transaction value and coded the customers in the top tercile (highest average transaction value) M = 3, and so forth. We coded the customers who made no repeat purchases R = F = M = 0.

In Table 2, we show the estimate of total CLV for the 28 resulting groups; we report the size of each RFM group in parentheses. Perhaps the most striking observation is the significant contribution of the "zero cell." Although each customer in that cell has a small CLV value (an average expected lifetime value of \$4.40 beyond Week 78 for someone who made his or her initial, and only, purchase at CDNOW in the first 12 weeks of the data set), this slight whisper of CLV becomes a loud roar when it is applied to such a large group of customers. This is an important substantive finding from our model. Many managers would assume that after a year and a half of inactivity, a customer has dropped out of the relationship with the firm, but these very light buyers collectively constitute approximately 5% of the total future value of the entire cohort, which is greater than most of the 27 other RFM cells.

Figure 14

DISTRIBUTION OF REPEATING CUSTOMERS IN THE 78-WEEK RECENCY/FREQUENCY SPACE

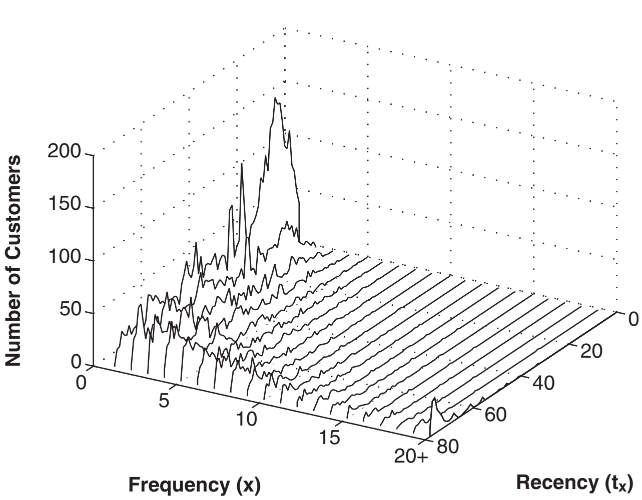


Table 2

TOTAL CLV BY RFM GROUP

		Recency			
	Frequency	*0*	*1*	*2*	*3*
M = 0	0	\$53,000 (12,054)			
M = 1	1		\$7,700 (1197)	\$9,900 (482)	\$1,800 (71)
	2		\$2,800 (382)	\$15,300 (488)	\$17,400 (419)
	3		\$300 (57)	\$12,500 (256)	\$52,900 (484)
M = 2	1		\$5,900 (650)	\$7,600 (264)	\$2,300 (68)
	2		\$3,600 (358)	\$26,500 (545)	\$25,800 (414)
	3		\$500 (86)	\$37,200 (478)	\$203,000 (972)
M = 3	1		\$11,300 (676)	\$19,700 (371)	\$3,700 (57)
	2		\$7,300 (329)	\$45,900 (504)	\$47,900 (396)
	3		\$1,000 (101)	\$62,700 (447)	\$414,900 (954)

Examining the other cells, we find clear evidence of the same patterns we discussed previously for the iso-value curves. For example, within the recency × frequency table associated with each level of the monetary value dimension, there is consistent evidence that high-frequency/low-recency customers are less valuable than those with lower frequency. Not surprisingly, the lower-right cell, which represents high levels on all three dimensions, has the greatest CLV; it shows an average net present value of $435 per customer. This represents approximately 38% of the future value of the entire cohort. For the cohort as a whole, the average CLV is approximately $47 per customer, making the full group of 23,560 customers worth slightly more than $1.1 million.

Table 3 reports the average CLV for each RFM tercile. We note that the greatest variability in CLV is on the recency dimension, followed closely by the frequency dimension. The least variation is on the monetary value dimension. This is consistent with the widely held view that recency is usually a more powerful discriminator than frequency or monetary value (thus, the framework is called RFM rather than, for example, FRM [Hughes 2000]).

SUMMARY AND CONCLUSION

Many recent articles have discussed the importance of determining CLV and understanding how it can be associated with observable behavioral characteristics, but despite this high level of interest in the topic, few have offered a carefully validated statistical model and specific CLV estimates for a large group of customers. From a methodological perspective, we offer several useful contributions to the CLV literature.

The key to our analysis is the use of a forward-looking model that features well-accepted behavioral assumptions rather than relying on only prior purchase data as a mirror for the future. Not only does the model help us translate prior behavior into likely future trends, but it also fills in the many sparse holes (and it smoothes out the random “blips” that are present in any observed data set), providing a cleaner image of these patterns. These advantages can be observed clearly in the iso-value curves that we discussed at length. A “data-driven” curve-fitting procedure would be hard-pressed to provide the same degree of diagnostic value as our model.

From the standpoint of managerial implementation, it is a significant benefit that the model inputs are nothing more than each customer’s RFM characteristics. Furthermore, our model also avoids the need to split the sample into two (or more) time periods for calibration purposes. All of the available data can be combined into a single sample to obtain the CLV forecasts. This uniform use of the data should lead to greater faith in the meaningfulness of these estimates and an improved ability to tie CLV to observed RFM characteristics.

Table 3

AVERAGE CLV BY RFM TERCILE

	Code		
	1	*2*	*3*
Recency	10	62	201
Frequency	18	50	205
Monetary value	31	81	160
N	3836	3835	3835

Beyond the performance of the model and its usefulness as a forecasting tool, we uncovered or confirmed several substantive observations:

- There is a highly nonlinear relationship between recency/frequency and future transactions. Standard scoring models (particularly those without solid underlying behavioral theories) would be unlikely to capture this complex relationship.
- The existence of the increasing frequency paradox is clearly visible in all of the iso-value curves. For low levels of recency, customers with greater frequency are likely to have lower future purchasing potential than customers with lower prior purchasing rates.
- The underlying process that drives monetary value per transaction appears to be stable over time and largely independent of recency and frequency. However, these patterns require further testing before they can be accepted as “empirical generalizations.” As such, any researcher applying our model to a new data set must test the validity of the assumption of independence between recency/frequency and monetary value. We believe that the analysis reported in the section “Assessing the Independence of the Monetary Value Assumption” and the conditional expectations of monetary value (Figure 8) are good ways to explore this matter.
- Despite the finding that monetary value and frequency are essentially independent, a reasonably strong regression-to-the-mean pattern creates the illusion that they are more tightly connected (Figure 13); that is, there is more regression-to-the-mean in $E(M|m_x, x)$ for customers with a small number of observed transactions (i.e., a low frequency) than for customers with a larger number of observed transactions (i.e., a high frequency). This is a general property of our submodel for dollar expenditure per transaction, which assumes that average transaction values are independent of the transaction process.
- Furthermore, the monetary value process does not conform well to the typical use of normal distributions for transaction amounts and intercustomer heterogeneity. The gamma-gamma model is a straightforward way to capture the high degree of skewness we observed in both of these distributions.
- A thorough analysis of the customer base requires careful consideration of the zero class (i.e., the customers who made no purchases during the observation period). This is often a very large group, so even though each member may have a small future lifetime value, collective profitability may still be quite substantial.
- We showed how iso-value curves can be created and used to identify customers with different purchase histories but similar CLVs.
- Finally, we emphasized the use of several validation tests (particularly the conditional expectations curves) that should be used in all CLV analyses regardless of the underlying model. Given the forward-looking nature of CLV, it is not enough to consider a model (and compare it with benchmarks) using only the calibration sample. Furthermore, given the individual-level nature of CLV, it is not enough to use tracking plots or other purely aggregate summaries to judge the performance of the model.

Although all of these contributions are managerially important, we consider our basic model only a first step toward a complete understanding of how best to understand, capture, and create additional value for a given group of customers. Natural extensions include (1) introducing marketing-mix variables into the model,[6] (2) adding an optimization layer to the model to help allocate resources most effectively, (3) relaxing the assumption of independence between the distribution of monetary value and the underlying transaction process, and (4) relaxing the assumption of constant contribution margin per transaction. With respect to the third point, this means we allow for a correlation between the ν parameter in the monetary value submodel and the λ parameter in the transaction model. This could be accommodated by replacing their respective (independent) gamma distributions with a bivariate Sarmanov distribution that has gamma marginals (Park and Fader 2004). Alternatively, this correlation between ν and λ could easily be accommodated by moving to a hierarchical Bayesian formulation of the basic model. With respect to the fourth point, our empirical analysis assumed a gross contribution margin of 30% because we have no information about the actual margins for CDNOW. In practice, we would expect to use customer-level (or at least segment-level) contribution margins. If we observe considerable intracustomer variation in contribution margin, we may wish to modify the monetary value component of the model (see the section "Adding Monetary Value") to model margin per transaction rather than dollar expenditure per transaction. We would need to replace the individual-level gamma distribution, the domain of which is nonnegative, with a skewed distribution defined over the domain of real numbers (to accommodate the transactions on which the company makes a loss).

We also recognize that our context of noncontractual purchasing is a limitation. It is true that, in general, RFM analyses are conducted in such a domain, but the use of CLV goes well beyond it. It is worthwhile to consider how similar models and concepts can be applied to subscriptions, financial services, and other types of business relationships. It is not immediately clear how the model components we used (Pareto/NBD for transaction incidence and gamma-gamma for transaction size) need to be changed to accommodate these other contexts, but we hope that researchers will continue to use stochastic models (rather than scoring models) to address these new situations.

Another limitation is that we examined only one cohort for CDNOW. This is a sizable group of customers but probably not a representative snapshot of the company's entire customer base. Managers need to run the model across multiple cohorts to obtain an accurate picture of the value of an entire customer base. In doing so, the definition of these cohorts becomes an important issue: Should we group customers simply by date of initial purchase (as we did here), or should we group them on the basis of geography, demographics, or mode of acquisition? The answers to these questions should not affect the development of the model per se, but they might have a large influence on how the model is implemented in practice.

[6]As Fader, Hardie, and Lee (2005b) note, such an exercise must be undertaken with extreme care. To the extent that customers have been targeted with different marketing incentives on the basis of their prior behaviors, econometric issues, such as endogeneity bias and sample selection bias, must receive serious attention.

With respect to acquisition, it would be desirable if the model could be used to make predictions for groups of potential buyers before the firm targets them for acquisition. It might be possible to "connect the dots" and extrapolate the model parameters from existing cohorts to new ones.

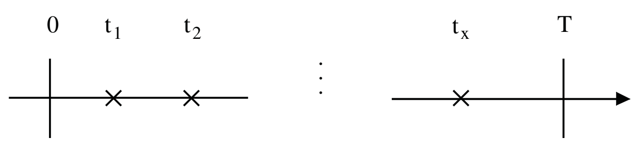
This latter point is an ambitious step. Further research must first replicate and extend the model for a variety of existing cohorts across a variety of firms. We expect that several of our substantive observations may become "empirical generalizations" related to CLV. We hope that other researchers and practitioners will confirm and extend some of our findings and uncover new behavioral patterns using the basic concepts and modeling platform we present herein.

APPENDIX

Our objective is to derive an expression for the present value of a customer's future transaction stream, conditional on observed purchasing history, as implied by the Pareto/NBD model. The Pareto/NBD model is based on the following assumptions:

1. While a customer is active, the number of transactions he or she makes follows a Poisson process with transaction rate λ. This is equivalent to assuming that the time between transactions is distributed exponentially with transaction rate λ.
2. Each customer has an unobserved lifetime of length τ (after which he or she is viewed as being inactive), which is distributed exponentially with dropout rate μ.
3. Heterogeneity in transaction rates across customers follows a gamma distribution with shape parameter r and scale parameter α.
4. Heterogeneity in dropout rates across customers follows a gamma distribution with shape parameter s and scale parameter β.
5. The transaction rate λ and the dropout rate μ vary independently across customers.

Assume that it is known when each of a customer's x transactions occurred during the period (0, T]; we denote these times by $t_1, t_2, \ldots, t_x$:



There are two possible ways this pattern of transactions could arise:

1. The customer is still active at the end of the observation period (i.e., $\tau > T$), in which case the individual-level likelihood function is simply the product of the (intertransaction time) exponential density functions and the associated survival function,

$$L(\lambda | t_1, t_2, \ldots, t_x, T, \tau > T)$$
$$= \lambda e^{-\lambda t_1} \lambda e^{-\lambda(t_2 - t_1)} \ldots \lambda e^{-\lambda(t_x - t_{x-1})} e^{-\lambda(T - t_x)}$$
$$= \lambda^x e^{-\lambda T}.$$

2. The customer became inactive at some time τ in the interval $(t_x, T]$, in which case the individual-level likelihood function is

$$L(\lambda | t_1, t_2, \ldots, t_x, T, \text{inactive at } \tau \in (t_x, T]) = \lambda^x e^{-\lambda \tau}.$$

Note that we do not require information on when each of the x transactions occurred; the only customer information we require is $(X = x, t_x, T)$. By definition, $t_x = 0$ when $x = 0$. In other words, recency (t_x) and frequency (x) are sufficient statistics.

Removing the conditioning on τ yields the following expression for the individual-level likelihood function:

$$\begin{aligned}
L(\lambda, \mu|X = x, t_x, T) &= L(\lambda|X = x, t_x, T, \tau > T)P(\tau > T|\mu) \\
&\quad + \int_{t_x}^{T}\{L(\lambda|X = x, t_x, T, \text{inactive at } \tau \in (t_x, T]) \\
&\quad \times f(\tau|\mu)\}d\tau \\
&= \lambda^x e^{-\lambda T}e^{-\mu T} + \lambda^x \int_{t_x}^{T} e^{-\lambda\tau}\mu e^{-\mu\tau}d\tau \\
&= \lambda^x e^{-(\lambda+\mu)T} + \frac{\mu\lambda^x}{\lambda+\mu}e^{-(\lambda+\mu)t_x} - \frac{\mu\lambda^x}{\lambda+\mu}e^{-(\lambda+\mu)T}.
\end{aligned}$$

It follows that the likelihood function for a randomly chosen person with purchase history $(X = x, t_x, T)$ is

$$\text{(A1)}\quad \begin{aligned}
L(r,\alpha,s,\beta|X = x,t_x,T) &= \int_0^\infty\int_0^\infty L(\lambda, \mu|X = x, t_x, T)g(\lambda|r,\alpha)g(\mu|s,\beta)d\lambda d\mu \\
&= \frac{\Gamma(r+x)\alpha^r\beta^s}{\Gamma(r)} \\
&\quad \times\left\{\frac{1}{(\alpha+T)^{r+x}(\beta+T)^s} + \left(\frac{s}{r+s+x}\right)A_0\right\},
\end{aligned}$$

where, for $\alpha \geq \beta$,

$$\begin{aligned}
A_0 &= \frac{1}{(\alpha+t_x)^{r+s+x}}\,{}_2F_1\left(r+s+x, s+1; r+s+x+1; \frac{\alpha-\beta}{\alpha+t_x}\right) \\
&\quad - \frac{1}{(\alpha+T)^{r+s+x}}\,{}_2F_1\left(r+s+x, s+1; r+s+x+1; \frac{\alpha-\beta}{\alpha+T}\right),
\end{aligned}$$

and for $\alpha \leq \beta$,

$$\begin{aligned}
A_0 &= \frac{1}{(\beta+t_x)^{r+s+x}}\,{}_2F_1\left(r+s+x, r+x; r+s+x+1; \frac{\beta-\alpha}{\beta+t_x}\right) \\
&\quad - \frac{1}{(\beta+T)^{r+s+x}}\,{}_2F_1\left(r+s+x, r+x; r+s+x+1; \frac{\beta-\alpha}{\beta+T}\right).
\end{aligned}$$

(For details of the derivation, see Fader and Hardie 2005.)

The four Pareto/NBD model parameters (r, α, s, β) can be estimated by the method of maximum likelihood in the following manner: Suppose we have a sample of N customers, where customer i had $X_i = x_i$ transactions in the period $(0, T_i]$, with the last transaction occurring at t_{x_i}. The sample log-likelihood function is given by

$$LL(r, \alpha, s, \beta) = \sum_{i=1}^{N}\ln\left[L(r, \alpha, s, \beta|X_i = x_i, t_{x_i}, T_i)\right].$$

This can be maximized with standard numerical optimization routines (see Fader, Hardie, and Lee 2005a).

The probability that a customer with purchase history $(X = x, t_x, T)$ is active at time T is the probability that the (unobserved) time at which he or she becomes inactive (τ) occurs after T. Referring back to our derivation of the individual-level likelihood function, the application of Bayes' theorem leads to

$$\text{(A2)}\quad \begin{aligned}
P(\tau > T|\lambda, \mu, X = x, t_x, T) &= \frac{L(\lambda|X = x, t_x, T, \tau > T)P(\tau > T|\mu)}{L(\lambda, \mu|X = x, t_x, T)} \\
&= \frac{\lambda^x e^{-(\lambda+\mu)T}}{L(\lambda, \mu|X = x, t_x, T)}.
\end{aligned}$$

We can now turn our attention to the derivation of DET as implied by the Pareto/NBD model. The general explicit formula for the computation of CLV is (Rosset et al. 2003)

$$CLV = \int_0^\infty v(t)S(t)d(t)dt,$$

where, for $t \geq 0$ (with $t = 0$ representing "now"), $v(t)$ is the customer's value at time t, $S(t)$ is the survivor function (i.e., the probability that the customer has remained active to at least time t), and $d(t)$ is a discount factor that reflects the present value of money received at time t. When we factor out the value of each transaction, $v(t)$ becomes the underlying transaction rate λ. It follows that, conditional on λ and μ, the discounted present value at time 0 of a customer's expected transaction stream over his or her lifetime with continuous compounding at rate of interest δ is

$$\text{(A3)}\quad \begin{aligned}
DET(\delta|\lambda,\mu) &= \int_0^\infty \lambda e^{-\mu t}e^{-\delta t}dt \\
&= \frac{\lambda}{\mu+\delta}.
\end{aligned}$$

As noted in any introductory finance textbook, an annual discount rate of $(100 \times d)\%$ is equivalent to a continuously compounded rate of $\delta = \ln(1 + d)$. If the data are recorded in time units such that there are k periods per year (k = 52 if the data are recorded in weekly units of time), the relevant continuously compounded rate is $\delta = \ln(1 + d)/k$.

Standing at time T, the DET for a customer with purchase history $(X = x, t_x, T)$ is the discounted present value of the customer's expected transaction stream over his or her lifetime (Equation A3) times the probability that the customer with this purchase history is active at time T (Equation A2):

$$\text{(A4)}\quad \begin{aligned}
&DET(\delta|\lambda, \mu, X = x, t_x, T) \\
&\quad = DET(\delta|\lambda, \mu)P(\tau > T|\lambda, \mu, X = x, t_x, T).
\end{aligned}$$

However, we do not observe λ and μ. Therefore, we compute $DET(\delta|X = x, t_x, T)$ for a randomly chosen person by taking the expectation of Equation A4 over the distribution of λ and μ, updated to account for the information $(X = x, t_x, T)$:

$$\text{(A5)}\quad \begin{aligned}
&DET(\delta|r, \alpha, s, \beta, X = x, t_x, T) \\
&\quad = \int_0^\infty\int_0^\infty\{DET(\delta|\lambda, \mu, X = x,t_x,T) \\
&\quad \times g(\lambda, \mu|r, \alpha, s, \beta, X = x, t_x, T)\}d\lambda d\mu.
\end{aligned}$$

According to Bayes' theorem, the joint posterior distribution of λ and μ is

$$\text{(A6)}\qquad g(\lambda, \mu|r, \alpha, s, \beta, X = x, t_x, T) = \frac{L(\lambda, \mu|X = x, t_x, T)g(\lambda|r, \alpha)g(\mu|s, \beta)}{L(r, \alpha, s, \beta|X = x, t_x, T)}.$$

Substituting Equations A2, A4, and A6 in Equation A5 leads to

$$\text{(A7)}\qquad DET(\delta|r, \alpha, s, \beta, X = x, t_x, T) = \int_0^\infty \int_0^\infty \frac{DET(\delta|\lambda, \mu)\lambda^x e^{-(\lambda + \mu)T} g(\lambda|r, \alpha)g(\mu|s, \beta)}{L(r, \alpha, s, \beta|X = x, t_x, T)} d\lambda d\mu.$$

Noting that Equation A3 can be written as a separable function of λ and μ, Equation A7 becomes

$$\text{(A8)}\quad DET(\delta|r, \alpha, s, \beta, X = x, t_x, T) = \frac{A \times B}{L(r, \alpha, s, \beta|X = x, t_x, T)},$$

where

$$\text{(A9)}\qquad A = \int_0^\infty \lambda^{x+1} e^{-\lambda T} g(\lambda|r, \alpha) d\lambda = \frac{\Gamma(r + x + 1)\alpha^r}{\Gamma(r)(\alpha + T)^{r + x + 1}}$$

and

$$B = \int_0^\infty \frac{1}{\mu + \delta} e^{-\mu T} g(\mu|s, \beta) d\mu = \frac{\beta^s}{\Gamma(s)} \int_0^\infty \frac{\mu^{s-1} e^{-\mu(\beta + T)}}{\mu + \delta} d\mu,$$

letting $z = \mu/\delta$ (which implies $\mu = \delta z$ and $d\mu = \delta dz$)

$$\text{(A10)}\qquad = \frac{\beta^s \delta^{s-1}}{\Gamma(s)} \int_0^\infty \frac{z^{s-1} e^{-z[\delta(\beta + T)]}}{z + 1} dz = \beta^s \delta^{s-1} \Psi[s, s; \delta(\beta + T)],$$

where $\Psi(\cdot)$ is the confluent hypergeometric function of the second kind (also known as the Tricomi function),[7] with integral representation

$$\Psi(a, c; Z) = \frac{1}{\Gamma(a)} \int_0^\infty e^{-zt} t^{a-1} (1 + t)^{c - a - 1} dt.$$

The confluent hypergeometric function of the second kind can be expressed in terms of two (regular) confluent hypergeometric functions,[8] or the incomplete gamma function when $a = c$.[9]

Substituting Equations A9 and A10 in Equation A8 leads to the following expression for DET, as implied by the Pareto/NBD model:

$$DET(\delta|r, \alpha, s, \beta, X = x, t_x, T) = \frac{\alpha^r \beta^s \delta^{s-1} \Gamma(r + x + 1)\Psi[s, s; \delta(\beta + T)]}{\Gamma(r)(\alpha + T)^{r + x + 1} L(r, \alpha, s, \beta|X = x, t_x, T)}.$$

[7]See http://functions.wolfram.com/HypergeometricFunctions/HypergeometricU/.

[8]See http://functions.wolfram.com/07.33.02.0001.01.

[9]See http://functions.wolfram.com/07.33.03.0003.01.

REFERENCES

Baesens Bart, Stijn Viaene, Dirk Van den Poel, Jan Vanthienen, and Guido Dedene (2002), "Using Bayesian Neural Networks for Repeat Purchase Modelling in Direct Marketing," *European Journal of Operational Research*, 138 (April), 191–211.

Bell, David, John Deighton, Werner J. Reinartz, Roland T. Rust, and Gordon Swartz (2002), "Seven Barriers to Customer Equity Management," *Journal of Service Research*, 5 (August), 77–86.

Berger, Paul D. and Nada I. Nasr (1998), "Customer Lifetime Value: Marketing Models and Applications," *Journal of Interactive Marketing*, 12 (Winter), 17–30.

Berry, Michael J.A. and Gordon S. Linoff (2004), *Data Mining Techniques*, 2d ed. Indianapolis, IN: John Wiley & Sons.

Blattberg, Robert C., Gary Getz, and Jacquelyn S. Thomas (2001), *Customer Equity*. Boston: Harvard Business School Press.

Bolton, Ruth N. (1998), "A Dynamic Model of the Duration of the Customer's Relationship with a Continuous Service Provider: The Role of Satisfaction," *Marketing Science*, 17 (1), 45–65.

Colombo, Richard and Weina Jiang (1999), "A Stochastic RFM Model," *Journal of Interactive Marketing*, 13 (Summer), 2–12.

Dwyer, F. Robert (1989), "Customer Lifetime Valuation to Support Marketing Decision Making," *Journal of Direct Marketing*, 3 (Autumn), 8–15.

Fader, Peter S. and Bruce G.S. Hardie (2001), "Forecasting Repeat Sales at CDNOW: A Case Study," *Interfaces*, 31 (May–June), S94–S107.

——— and ——— (2005), "A Note on Deriving the Pareto/NBD Model and Related Expressions," (accessed July 31, 2005), [available at http://brucehardie.com/notes/009/].

———, ———, and Ka Lok Lee (2005a), "A Note on Implementing the Pareto/NBD Model in MATLAB," (accessed June 23, 2005), [available at http://brucehardie.com/notes/008/].

———, ———, and ——— (2005b), "'Counting Your Customers' the Easy Way: An Alternative to the Pareto/NBD Model," *Marketing Science*, 24 (Spring), 275–84.

Hughes, Arthur M. (2000), *Strategic Database Marketing*, 2d ed. New York: McGraw-Hill.

Malthouse, Edward C. (2003), "Scoring Models," in *Kellogg on Integrated Marketing*, Dawn Iacobucci and Bobby Calder, eds. Hoboken, NJ: John Wiley & Sons, 227–49.

——— and Robert C. Blattberg (2005), "Can We Predict Customer Lifetime Value?" *Journal of Interactive Marketing*, 19 (Winter), 2–16.

Morrison, Donald G. and Jorge Silva-Risso (1995), "A Latent Look at Empirical Generalizations," *Marketing Science*, 14 (3), G61–G70.

Mulhern, Francis J. (1999), "Customer Profitability Analysis: Measurement, Concentrations, and Research Directions," *Journal of Interactive Marketing*, 13 (Winter), 25–40.

Park, Young-Hoon and Peter S. Fader. (2004), "Modeling Browsing Behavior at Multiple Websites," *Marketing Science*, 23 (Summer), 280–303.

Parr Rud, Olivia (2001), *Data Mining Cookbook*. New York: John Wiley & Sons.

Pfeifer, Phillip E. and Robert L. Carraway (2000), "Modeling Customer Relationships as Markov Chains," *Journal of Interactive Marketing*, 14 (Spring), 43–55.

———, Mark E. Haskins, and Robert M. Conroy (2005), "Customer Lifetime Value, Customer Profitability, and the Treatment of Acquisition Spending," *Journal of Managerial Issues*, 17 (1), 11–25.

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

——— and ——— (2003), "The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration," *Journal of Marketing*, 67 (January), 77–99.

Rosset, Saharon, Einat Neumann, Uri Eick, and Nurit Vatnik (2003), "Customer Lifetime Value Models for Decision Support," *Data Mining and Knowledge Discovery*, 7 (July), 321–39.

Schmittlein, David C., Lee G. Cooper, and Donald G. Morrison (1993), "Truth in Concentration in the Land of (80/20) Laws," *Marketing Science*, 12 (Spring), 167–83.

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

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