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

Peter S. Fader
Bruce G. S. Hardie
Ka Lok Lee[1]

July 2004
Revised December 2004
Revised February 2005

[1]Peter S. Fader is the Frances and Pei-Yuan Chia Professor of Marketing at the Wharton School of the University of Pennsylvania (address: 749 Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104-6340; phone: 215.898.1132; email: faderp@wharton.upenn.edu; web: www.petefader.com). Bruce G. S. Hardie is Associate Professor of Marketing, London Business School (email: bhardie@london.edu; web: www.brucehardie.com). Ka Lok Lee is Market Research Analyst, Catalina Health Resource (email: kaloklee@alumni.upenn.edu). The authors thank Albert Bemmaor, Paul Farris, Phil Pfeifer, three anonymous referees, and seminar participants at the Yale SOM and University of Auckland for their helpful comments. The second author acknowledges the support of the London Business School Centre for Marketing.

Abstract

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

We present a new model that links the well-known RFM (recency, frequency, monetary value) paradigm with customer lifetime value (CLV). While previous researchers have connected the two conceptually, none has presented a formal model that requires nothing more than RFM inputs to make specific lifetime value projections for a set of customers. Key to this analysis is the notion of "iso-value" curves, which enable us to group together individual customers who have different behavioral histories but similar future valuations. Iso-value curves make it easy to visualize and summarize the main interactions and tradeoffs among the RFM measures and CLV. Our stochastic model, featuring a Pareto/NBD framework to capture the flow of transactions over time and a gamma-gamma sub-model for dollars per transaction, reveals a number of subtle but important non-linear associations that would be missed by relying on observed data alone. We conduct a number of holdout tests to demonstrate the validity of the model's underlying components, and then focus on our iso-value analysis to estimate the net present value for a large group of customers of the online music site, CDNOW. We summarize a number of substantive insights and point out a set of broader issues and opportunities in applying such a model in actual practice.

Keywords: customer lifetime value, CLV, RFM, customer base analysis, Pareto/NBD

1 Introduction

The move towards 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 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 cashflows associated with a customer (Pfeifer et al. 2005). Because of the challenges associated with forecasting future revenue streams, most empirical research on "lifetime value" has actually computed customer profitability based solely on customers' past behavior. But in order to be true to the notion of CLV, our measures should look to the future, not the past. A significant barrier has been our 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).

A number of researchers and consultants have developed "scoring models" (i.e., regression-type models) that attempt to predict a customer's future behavior. (See, for example, 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 a customer's past behavior are key predictors of their future behavior. Drawing on the direct marketing literature, it is common practice to summarize a customer's past behavior in terms of their "RFM" characteristics: *recency* (time of most recent purchase), *frequency* (number of past purchases), and *monetary value* (average purchase amount per transaction).

But there are several problems with these models, especially when seeking 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; we also need to predict behavior 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 we were able to create predictions of future purchasing behavior simply using period 1 data alone. 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 fact 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 we utilize the observed data from one period to make predictions of future behavior.

The way to overcome these general problems is to develop a formal model of buyer behavior, 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 an individual'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 measures of recency, frequency, and monetary value 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 in order to implement our model.

Before conveying our analytical results, we offer a brief exploratory analysis to set the stage for our model development. Let us consider the purchasing of the cohort of 23,570 individuals who made their first-ever purchase at CDNOW in the first quarter of 1997. We have data covering their initial and subsequent (i.e., repeat) purchases up to the end of June 1998. (See

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

Fader and Hardie 2001 for further details about this dataset.) For presentational simplicity, we initially examine the relationship between future purchasing and recency/frequency alone; we will introduce monetary value later in the paper. We first split the 78-week dataset 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 are presented in Figure 1.

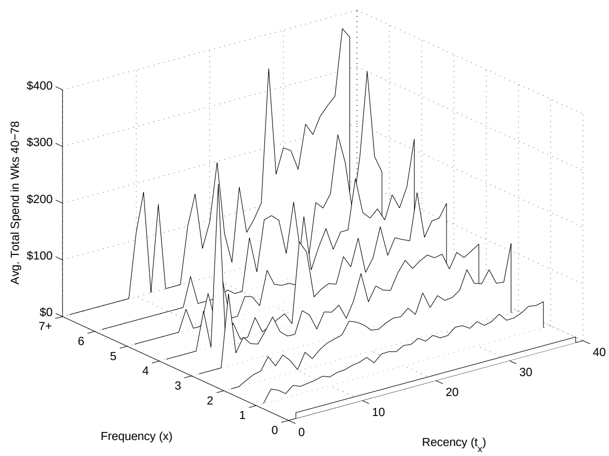


Figure 1: Average Total Spend in Weeks 40–78 by Recency and Frequency in Weeks 1–39

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 the figure). But despite the large number of customers used to generate this graph, it is still very sparse and therefore somewhat untrustworthy. A number of the "valleys" in Figure 1 are simply due to the absence of any observations for particular combinations of recency and frequency. It is important for us to abstract away from the observed data and develop a formal model in order to "fill in the blanks."

An alternative view of the same relationship is shown in Figure 2. This contour plot, or "30,000 ft view" of Figure 1, is an example of an "iso-value" plot; each curve in the graph links together customers with equivalent future value despite differences in their past behavior.

[2] We have removed the purchasing data for ten buyers who purchased over \$4000 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. This initial exploratory analysis is therefore based on the purchasing of 23,560 customers.

Again, it is easy to see how higher recency and higher frequency are correlated with greater future purchasing, but the jagged nature of these curves highlights the dangers of relying solely on observed data in the absence of a model.

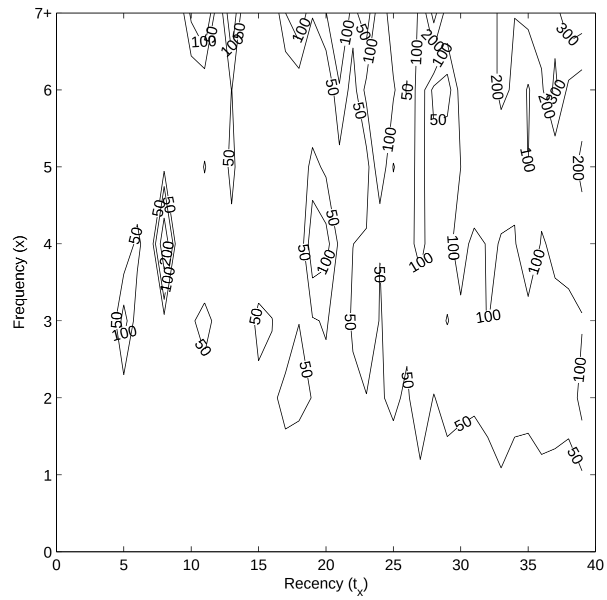


Figure 2: Contour Plot of Average Week 40–78 Total Spend by Recency and Frequency

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).

But despite these limitations in attempting to use the kind of data summaries shown 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 seen in Figure 2 into the stylized version shown 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 allow us to get a better understanding of these relationships, 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. (An obvious parallel is Schmittlein et al.'s (1993) work on "80/20 rules" in purchasing, where a model is used 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

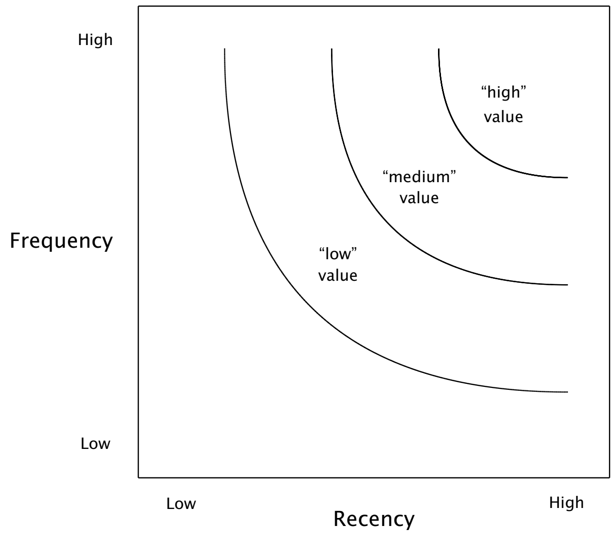


Figure 3: Hypothesized Relationship Between Recency/Frequency and CLV

evaluate the CLV for an entire customer base. It can help guide managerial decision making and provide accurate quantitative benchmarks to better gauge the "return on investment" for programs that companies use to develop and manage their portfolio of customers.

In the next section we describe the model that underpins our effort to link RFM and CLV, one for which recency, frequency, and monetary value are sufficient statistics for an individual customer's purchasing history. After briefly discussing the model results and some hold-out validations to assess its efficacy, we turn to a detailed discussion of the iso-value curves. Following that, we use our new model to compute CLV for the entire cohort of CDNOW customers discussed earlier. We close with a brief summary of our work and a discussion of promising future research directions.

2 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 can therefore be trusted as the basis for any CLV estimates. Ideally such a model would generate these estimates using only simple summary statistics (such as RFM) without requiring more detailed information about each customer's purchasing history.

In developing our model, we will assume that monetary value is independent of the underlying transaction process. While this may seem counterintuitive to some (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 $\times$ contribution margin) can be factored out and we can focus on forecasting the "flow" of future transactions (discounted to yield a present value). This number of discounted expected transactions (DET) can then be rescaled by a value "multiplier" to yield an overall estimate of lifetime value:

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

We first develop our sub-model for DET alone, then introduce a separate sub-model for expected revenue per transaction. Despite the fact that we assume independence for these two processes, we will still see some interesting relationships (particularly between frequency and monetary value) that we will explore later.

The Pareto/NBD model developed by Schmittlein, Morrison, and Colombo (1987, hereafter SMC) has proven itself to be a popular and powerful model to explain the flow of transactions in a noncontractual setting. Excellent illustrations are provided by Reinartz and Kumar (2000, 2003). This model is based on the following general assumptions regarding the repeat buying process:

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

ii. While active, customers can place orders whenever they want to. The number of orders placed by a customer in any given time period (e.g., week or month) appears to vary randomly about his underlying average rate.

iii. Customers (while active) vary in their underlying average purchase rates.

iv. 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: a long hiatus does not necessarily mean that the customer has actually become inactive. There is no way for an outside observer to know for sure (hence the need for a model to make a "best guess" about this process).

v. Customers become inactive for any number of reasons and so the unobserved time at which the customer becomes inactive appears to have a random component.

vi. This inclination for customers to "dropout" of their relationship with the firm is heterogeneous. In other words, some customers are expected to become inactive much sooner than others. Some may remain active for many years — well beyond the length of any conceivable dataset.

vii. Purchase rates (while active) and "dropout" rates vary independently across customers.

Translating these general assumptions into specific mathematical assumptions gives us the Pareto/NBD model.

The only customer-level information required by this model 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.

SMC derive expressions for a number of 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.

- $P(\text{"active"} \mid X = x, t_x, T)$, the probability that an individual with observed behavior $(X = x, t_x, T)$ is still "active" at time T.

- $E(Y(t) \mid X = x, t_x, T)$, the expected number of transactions in the future period $(T, T + t]$ for an individual 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 discounted expected transactions for a customer with observed behavior $(X = x, t_x, T)$ as

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

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 faced by anyone performing CLV calculations (Blattberg et al. 2001): (i) what time horizon to use in projecting sales?, and (ii) what time periods to measure (e.g., year, quarter)? Furthermore, it 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 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 following expression for DET as implied by the Pareto/NBD model:

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

where r, α, s, 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, an expression for which is given in (A1). The derivation of this expression for DET is a new analytical result and is central to our CLV estimation.

2.1 Adding Monetary Value

Up to 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, we are more interested in the total dollar value across these transactions (from which a customer's profitability can then be calculated). For this we need a separate sub-model for dollar expenditure per transaction.

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

i. The dollar value of a customer's given transaction varies randomly around his average transaction value

ii. Average transaction values vary across customers but do not vary over time for any given individual

iii. The distribution of average transaction values across customers is independent of the transaction process.

In some situations (e.g., purchasing of a product that a customer can hold in inventory), some analysts may expect to see a relationship between transaction timing and quantity/value. However, for the situation at hand, we will assume that recency/frequency and monetary value are independent. (We explore the validity of this assumption in Section 2.2 below.)

Schmittlein and Peterson (1994) assume that the random purchasing around each individual's mean is characterized by a normal distribution, and that the average transactions 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 will be 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 completing 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 individuals who made at least one repeat purchase in weeks 1–39. Seeing the big differences in the mean, median, and mode, it is clear that the distribution of *observed* individual means is highly right-skewed. This suggests that the *unobserved* heterogeneity in the individual-level means cannot be characterized by a normal distribution.

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

Table 1: Summary of Average Repeat Transaction Value per Customer (Weeks 1–39)

We could modify Schmittlein and Peterson's 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 individual's mean) with one whose domain is nonnegative. An obvious choice would be the lognormal distribution. However, as 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 his *true* mean transaction rate, a quantity that is central to our CLV calculations). We therefore turn to the gamma distribution, which has similar properties to those of the lognormal (albeit with a slightly thinner tail), adapting the gamma-gamma model originally proposed by Colombo and Jiang (1999).

To help understand the role of the monetary value sub-model, 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 instance, suppose that the mean expediture across all customers across all transactions is \$35, but customer A has made one repeat purchase totalling \$100. What value of $E(M)$ should we use for our future projection of customer A? Should we assume that $E(M) = m_1 = \$100$ or should we "debias" our estimate down towards the population mean? As we see more and more repeat purchases for each customer, we would expect that the observed value of m_x will become increasingly accurate as an estimate of their 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 $m_x = \sum_{i=1}^{x} z_i/x$ is an imperfect estimate of $E(M)$. Our goal is to make inferences about $E(M)$ given m_x, which we denote as $E(M|m_x, x)$. Clearly $m_x \to E(M)$ as $x \to \infty$, but this could be a slow process.

We assume that the Z_i are iid gamma variables with shape parameter p and scale parameter ν. Utilizing two standard relationships involving the gamma distribution,

i. the sum of x iid gamma(p, ν) random variables is distributed gamma with shape parameter px and scale parameter ν, and

ii. 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. (The parameter p is assumed to be 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 gives us the following marginal distribution for m_x:

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

In order to arrive at an expression for our desired quantity, $E(M|m_x, x)$, we employ 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 see less weight being placed on the population mean and more weight on the observed customer-level average of m_x. This "regression-to-the-mean" phenomenon will be illustrated in more detail when we revisit the monetary value sub-model later in the paper.

2.2 Assessing the Independence of 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 individuals who made at least one repeat purchase in weeks 1–39 (out of a sample of 2357 customers), we find that the simple correlation between average transaction value and the number of transactions is 0.11. 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 0.06 ($p = 0.08$).

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. While we can recognize the slight positive correlation, it is clear that the variation within each number-of-transactions group dominates the between-group variation.

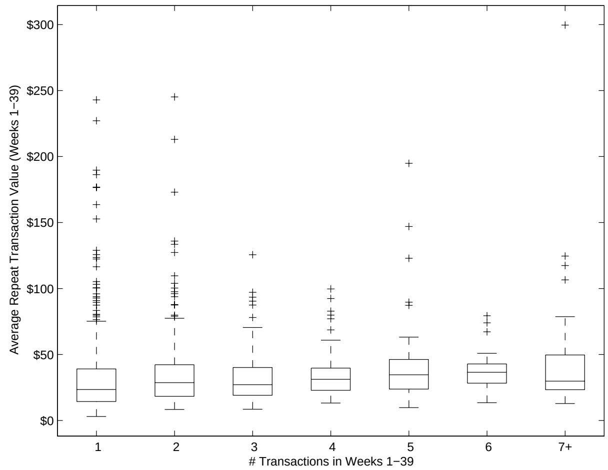


Figure 4: The Distribution of Average Transaction Value by Number of Transactions

So while there is a slight positive correlation between average transaction value and the number of transactions, we do not feel that it represents a substantial violation of our independence assumption. In the next section, we will provide some additional evidence that this modest relationship has no discernable impact on the performance of our model in a holdout setting (see the discussion surrounding Figure 8).

3 Model Validation

Before we use the expressions developed above to explore the relationship between RFM and CLV for the case of CDNOW, it is important to verify that the Pareto/NBD sub-model for transactions and the gamma-gamma sub-model for monetary value each provides accurate predictions of future buying behavior. A careful holdout validation like this is missing from virtually all CLV-related papers but must be conducted before making 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 will utilize all 78 weeks of data for our CLV analysis.)

First, 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 conducted by Fader et al. (2005a).

The maximum likelihood estimates of the model parameters are $\hat{r} = 0.55$, $\hat{\alpha} = 10.58$, $\hat{s} = 0.61$, 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, under-forecasting by less than 2% at week 78.

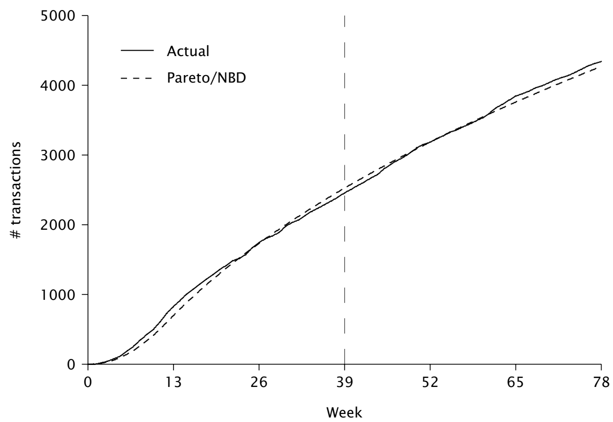


Figure 5: Tracking Cumulative Repeat Transactions

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 past 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 that took place in the forecast period, broken down by the number of repeat purchases in the first 39 weeks. (Note that for each frequency level (x), we are averaging over customers of differing recency (i.e., different values of t_x).

Our ability to closely track the sales data in a holdout period (both at the aggregate and individual level) gives us confidence in the Pareto/NBD model as the basis for our upcoming CLV calculations.

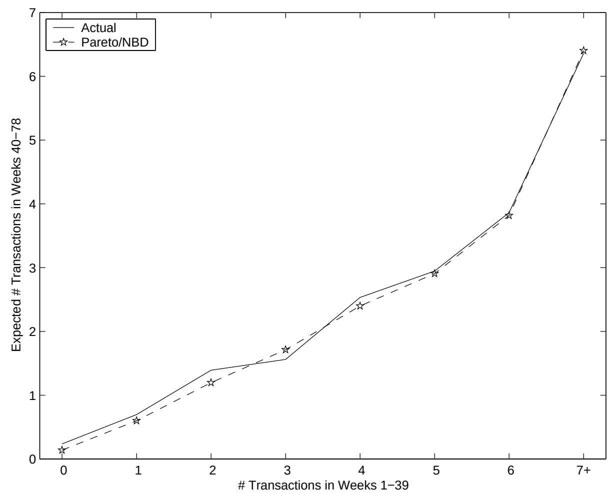


Figure 6: Conditional Expectations of Purchasing (Weeks 40–78)

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 nine cents. To visualize the fit of the model, we compute the implied distribution of average transaction value across individuals using (3) and compare it to a nonparametric density of the observed average transaction values in Figure 7.

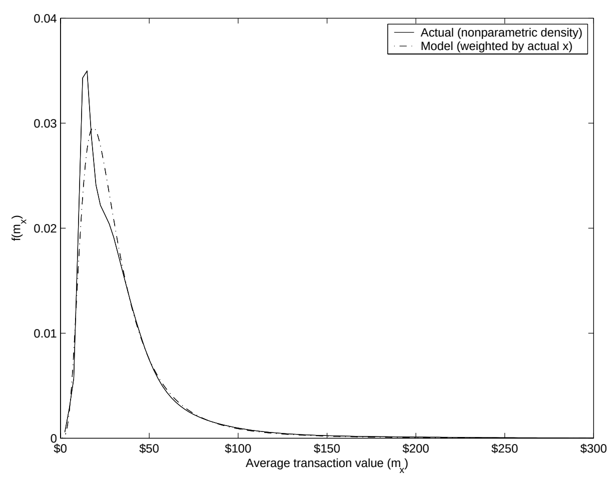


Figure 7: Observed vs. Theoretical Distribution of Average Transaction Value Across Customers

The fit is reasonable. However, the theoretical mode of \$19 is higher 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 paper we choose to forgo such tweaks in order 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 in the forecast period (Weeks 40–78) of the dataset. For each customer, we compute the expected average transaction value conditioned on his calibration-period frequency and monetary value using (4). We compute two sets of conditional expectations of forecast-period monetary value: the first is obtained by multiplying each individual's conditional expectation of per-transaction value by his *actual* number of forecast-period transactions; the second is obtained by multiplying the individual's conditional expectation of per-transaction value by his *predicted* number of forecast-period transactions, conditional on his calibration-period recency and frequency (i.e., the numbers summarized in Figure 6). These different projections let us test each sub-component 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 are averaging over customers of differing recency (values of t_x) and calibration-period monetary value).

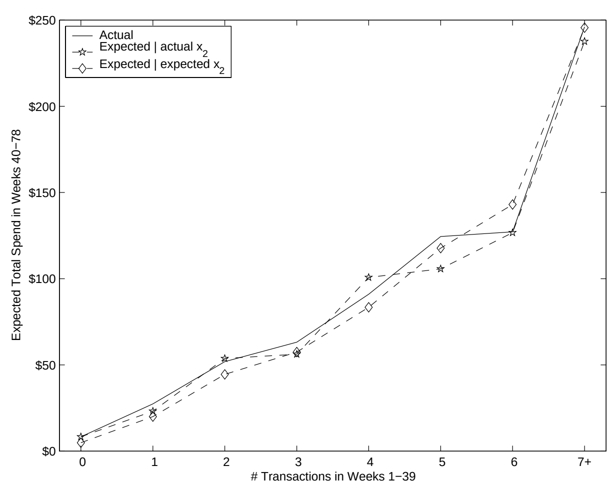


Figure 8: Conditional Expectations of Monetary Value

Comparing the "expected | actual x_2" numbers with the actual numbers can be viewed as a clean test of the monetary-value sub-model. In comparing these predictions with the actual numbers, there does not appear to be any particular bias that would lead us to question the assumption of stationarity in the average buying rates problem or the assumption that average transaction value is independent of frequency. Comparing the "expected | expected x_2" numbers with the actual numbers can be viewed as a test of the combined Pareto/NBD + gamma-gamma models. The combined models provide good conditional predictions of expected total monetary value in the forecast period, which gives us confidence in using this model to make overall predictions of lifetime value outside the range of the observed data.

4 Creating and Analyzing Iso-value Curves

Now that we have demonstrated the substantial validity of the Pareto/NBD sub-model for transactions and the gamma-gamma sub-model for monetary value, we can use all 78 weeks of data to make predictions of CLV out into the future. As noted earlier, 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; 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.

First we re-estimate the models using all 78 weeks of data, and as a final validity check we compare this "full" model to the 39-week version used in the previous section. Fortunately the model remains very 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 function, we notice only a small decrease in model fit (from -9608 to -9595). For the gamma-gamma model, the degree of stability is truly remarkable: The 78-week parameters provide a 39-week LL of -4661, compared to the optimal 39-week LL of -4659. This offers strong support for our assumption that the sub-model governing monetary value is stable over time.

We begin our iso-value analysis by focusing first on the relationship between discounted expected transactions (DET) and recency/frequency. We then reintroduce monetary value to complete the picture.

To obtain the iso-value curves for DET, we evaluate (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 $\delta = 0.0027$. The CLV estimates are presented as a "waterfall" plot in Figure 9.

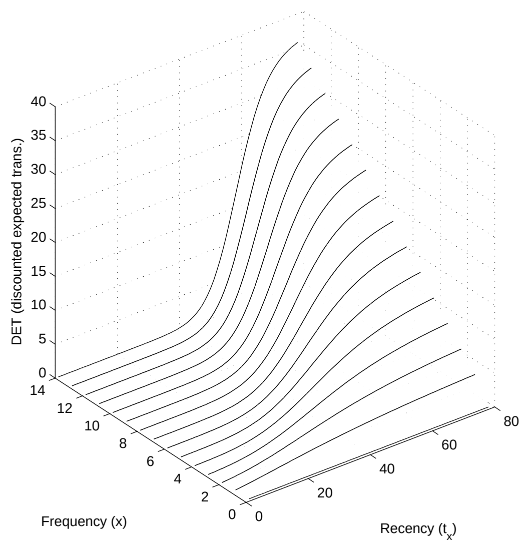


Figure 9: DET as a Function of Recency and Frequency

With the exception of $x = 0$,[4] DET is an increasing function of recency. Note, however, 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 bigger role in determining CLV than for an infrequent past purchaser.

Further light on this complex relationship can be seen in the iso-value curves shown in Figure 10. For the high value customers (i.e., upper right of Figure 10), the iso-value curves reflect the basic shape hypothesized in Figure 3. But as we move toward the lower-value regions, we observe that the iso-value lines start to bend backwards. At first thought, this seems highly counterintuitive: someone with frequency of $x = 7$ and recency of $t_x = 35$ has an approximate

[3]For both computational and presentational simplicity, these calculations are performed with $T = 77.86$ (i.e., for a hypothetical customer who made his first-ever purchase at CDNOW on January 1, 1997.

[4]For the case of $x = 0$ (i.e., no repeat purchases), recency has no meaning. Following SMC, we let $t_x = 0$ for such cases.

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.

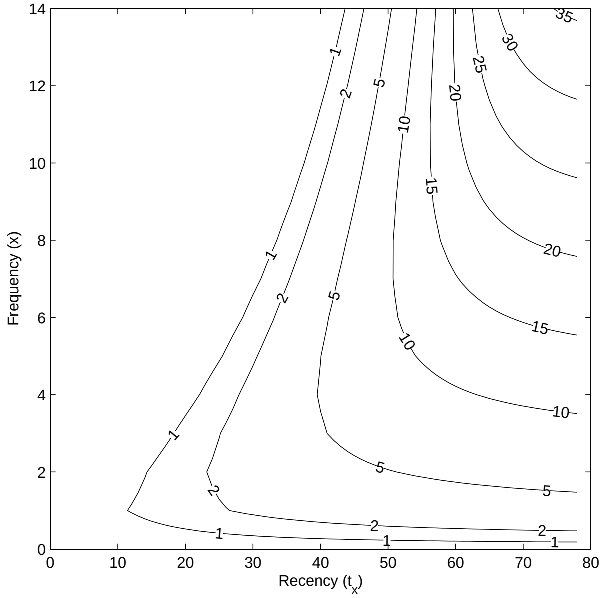


Figure 10: Iso-value Representation of DET

To resolve this apparent paradox, consider the two hypothetical customers shown in Figure 11. If we knew for sure that both customers were still active in week 78, we would expect customer B to have a greater value of DET, given his higher number of past purchases. But the pattern of purchases strongly suggests that customer B is no longer active. Customer A, on the other hand, has a lower underlying purchase rate, so it is more likely that he is still active in week 78. The net effect is that the DET for customer A is 4.6, versus 1.9 for customer B.

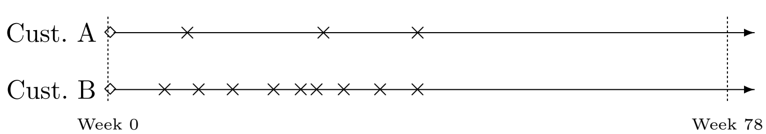


Figure 11: The "Increasing Frequency" Paradox

Therefore, upon further reflection, the existence of backward bending iso-value curves makes sense. In our dataset we see a number of customers with purchasing histories qualitatively similar to that 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 seen in the actual data (Figure 2), despite the fact that we used data for 23,560 customers to create it, makes it hard to properly visualize these vital relationships. Furthermore, the observed shape of the relationship does not fully reflect the true relationship (just as Schmittlein et al.'s (1993) work on "80/20 rules" in purchasing demonstrates that what we see in 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 instead of 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 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 each customer based on their purchase history (i.e., recency, frequency and monetary value) to date? Given our assumption of independence between the transaction and monetary-value processes, it is tempting to simply multiply the DET (as predicted given recency and frequency) by the individual's observed average monetary value (m_x). But this would ignore the "regression-to-the-mean" phenomenon discussed earlier and formalized in (4). Instead we need to take into account the number of transactions made by each customer, and use that information about their transaction history to come up with a weighted average between their past purchasing and the overall population tendencies. This is illustrated in Figure 12 for three different values of average observed transaction values (m_x): \$20, \$35, \$50.

If no repeat purchases are observed in weeks 1–78, our best guess is that the individual's average transaction value in the future is simply the mean of the overall transaction value distribution. This mean for full 78-week period is \$36 and identified by the circle in Figure 12. Our prediction of lifetime value for a customer with zero repeat purchases would therefore be

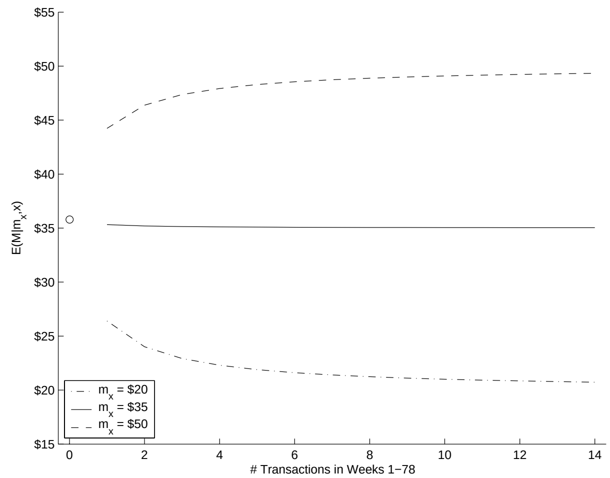


Figure 12: DET Multipliers as a Function of Frequency and Average Transaction Value

36 times the DET value for $x = 0$ as reported in Figure 9 (times the gross margin).

With one observed repeat purchase, the best guess for $E(M)$ moves towards the observed value of m_x (i.e., the amount of that single transaction) but not all the way there. Not until the customer has made 7–8 transactions can we trust the observed value of m_x to serve as a reasonably accurate estimate of his 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 utilized by other researchers (e.g., Reinartz and Kumar 2000, 2003). Choosing a different margin will obviously change the values reflected in the subsequent iso-value curves, but should not affect any of the main patterns seen 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 (2) and (4) in (1) along with our assumed margin of 30%. Given an assumed annual discount rate of 15% and 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 plug-in the observed values for RFM (t_x, x, and m_x, respectively) to get an expected CLV for the customers who share those behavioral characteristics. The ability to obtain (presumably) accurate estimates for CLV from these basic inputs is, by itself, a significant contribution from this work.

From a graphical standpoint, a complete view of this relationship would require us to move from the 3-dimensional plots seen earlier (Figures 1 and 9) to a 4-dimensional representation. While 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 used before, but 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 used earlier in Figure 12. At first glance, these waterfall plots show some meaningful similarities to each other and to the earlier plots for DET alone (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 side of each graph reflects the "increasing frequency" paradox discussed earlier.

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

Another way of understanding and appreciating the usefulness of our CLV estimation method is to combine the model-driven RFM-CLV relationship with the actual RFM patterns seen in our dataset. In Figure 14 we show the distribution in the recency/frequency space for the 11,506 customers (out of 23,560) who made at least one repeat purchase at CDNOW in the 78-week observation period. Note that the directions of the recency and frequency axes have

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

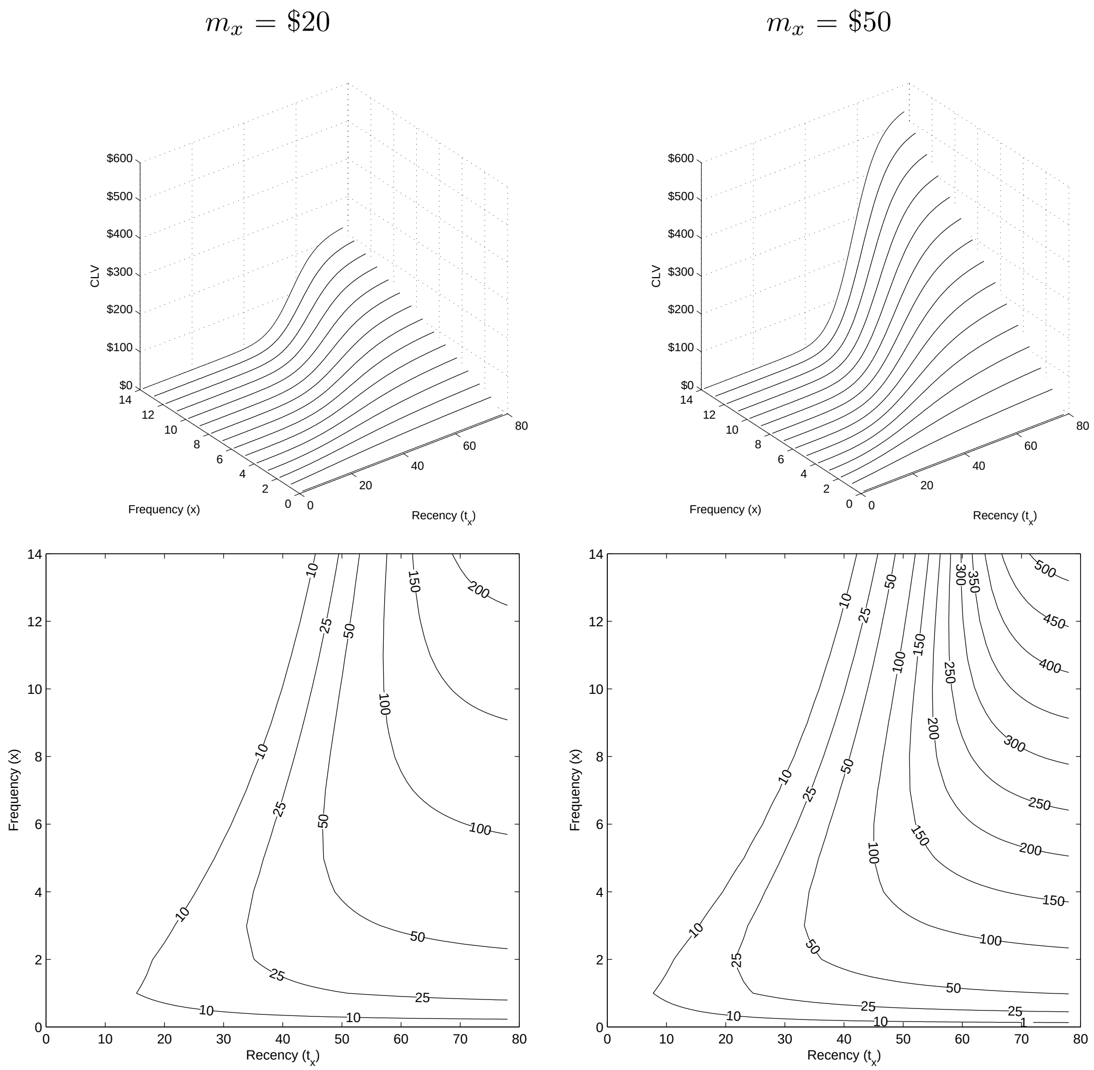


Figure 13: CLV as a Function of R & F for Average Transaction Values of \$20 and \$50

been reversed (for this figure only) to make it possible to see the patterns here. This figure shows that the majority of repeating customers had only one repeat transaction; noting 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.

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

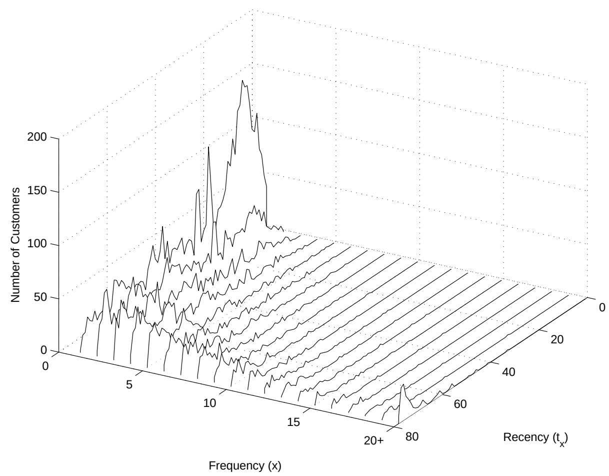


Figure 14: Distribution of Repeating Customers in the 78-week Recency/Frequency Space

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

In Table 2 we show the estimate of total CLV for the 28 resulting groups; the size of each RFM group is reported in parentheses. Perhaps the most striking observation is the significant contribution of the "zero cell." Even though each customer in that cell has a very small CLV value (an average expected lifetime value of \$4.40 beyond week 78 for someone who made their initial — and only — purchase at CDNOW in the first twelve weeks of the dataset). But this slight whisper of CLV becomes a loud roar when 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 their relationship with the firm. But these very light buyers collectively constitute almost 5% of the total future value of the entire cohort — larger than most of the 27 other RFM cells.

		Recency			
	Frequency	0	1	2	3
M=0	0	$53,000 (1,197)			
M=1	1		$7,700 (1,197)	$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)

Table 2: Total CLV by RFM Group

Looking at those other cells, we see clear evidence of the same patterns discussed earlier for the iso-value curves. For instance within the R×F table associated with each level of the M 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, representing high levels on all three dimensions, has the greatest CLV, with an average net present value of $435 per customer. This represents nearly 38% of the future value of the entire cohort. For the cohort as a whole, the average CLV is about $47 per customer, making the full group of 23,560 customers worth just over $1.1 million.

Table 3 reports the average CLV for each recency, frequency, and monetary-value tercile. We note that there is greatest variability in CLV on the recency dimension, closely followed by the frequency dimension. There is least variation 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. (This is why the framework is called "RFM" instead of "FRM", etc. (Hughes 2000).)

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

Table 3: Average CLV by Recency, Frequency, and Monetary-Value Tercile

5 Summary and Conclusions

Many recent papers 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, very few papers have offered a carefully validated statistical model and specific CLV estimates for a large group of customers. From a methodological perspective, these are useful contributions to the CLV literature.

The key to this analysis is the use of a forward-looking model, featuring well-accepted behavioral assumptions, instead of relying on past purchase data alone as a mirror for the future. Not only does the model help us translate past behavior into likely future trends, but it also fills in the many sparse holes (and smooths out the random "blips" present in any observed dataset) to give us a cleaner image of these patterns. These advantages can be seen 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, the fact that the model inputs are nothing more than each customer's recency, frequency, and monetary value is a significant benefit. 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 give us greater faith in the meaningfulness of these estimates, as well as an improved ability to tie CLV to observed RFM characteristics.

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

- 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, as seen clearly in all of the iso-value curves. For low levels of recency, customers with higher frequency are likely to have lower future purchasing potential than customers with lower past purchasing rates.

- The underlying process that drives monetary value per transaction appears to be stable over time and largely independent of recency and frequency. Of course these patterns require much more further testing before they can be accepted as "empirical generalizations." As such, any researcher applying our model to a new dataset must test the validity of the assumption of independence between recency/frequency and monetary value. We feel that the analysis reported in Section 2.2 and the conditional expectations of monetary value (Figure 8) are good ways of exploring this matter.

- Despite the finding that monetary value and frequency are basically independent, a reasonably strong "regression-to-the-mean" pattern creates the illusion that they seem to be 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 those customers with a larger number of observed transactions (i.e., a high frequency). This is a general property of our sub-model 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 very well to the typical use of normal distributions for transaction amounts as well as inter-customer heterogeneity. The gamma-gamma model is a straightforward way to capture the high degree of skewness seen in both of these distributions.

- A thorough analysis of the customer base requires careful consideration of the "zero class," i.e., those customers who made no purchases during the observation period. This is often a very large group, so despite the fact that each member may have a very small future lifetime value, their 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 a number of validation tests (particularly the conditional expectations curves) that should be employed in all CLV analyses regardless of the underlying model. Given the forward-looking nature of CLV, it is not enough to look at a model (and compare it to benchmarks) using the calibration sample alone. 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.

While all of these contributions are managerially important, we consider our basic model to be only a first step towards a complete understanding of how best to understand, capture, and create additional value for a given group of customers. Natural extensions include: (i) introducing marketing mix variables into the model,[6] (ii) adding an optimization layer to the model to help allocate resources most effectively, (iii) relaxing the assumption of independence between the distribution of monetary value and the underlying transaction process, and (iv) 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 sub-model 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 last point, our empirical analysis assumed a gross contribution margin of 30% as 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 intra-customer variation in contribution margin, we may wish to

[6] As noted by Fader et al. (2005a), 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 past behaviors, econometric issues such as endogeneity bias and sample selection bias must receive serious attention.

modify the monetary value component of the model (section 2.1) to model margin per transaction instead of dollar expenditure per transaction. We would have to replace the individual-level gamma distribution, whose domain is nonnegative, with a skewed distribution defined over the domain of real numbers (to accommodate those transactions on which the company makes a loss).

We also recognize that the context here — noncontractual purchasing — is a limitation. It is true that RFM analyses are generally 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 used here (Pareto/NBD for transaction incidence, gamma-gamma for transaction size) need to be changed to accommodate these other contexts, but we hope that future researchers will continue to use stochastic models (instead of scoring models) to address these new situations.

Another limitation is that we looked at only one cohort for CDNOW. This is a sizeable group of customers, but probably not a representative snapshot of their entire customer base. Managers need to run the model across multiple cohorts in order 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 based on geography, demographics, or mode of acquisition instead? 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.

And thinking of acquisition, it would be very desirable if the model could be used to make predictions for groups of potential buyers *before* they are targeted for acquisition by the firm. It might be possible to play "connect the dots" and extrapolate the model parameters from existing cohorts to new ones.

But this latter point is an ambitious step. First we need to replicate and extend the model for a variety of existing cohorts across a variety of firms. Our expectation is that several of our substantive observations may become "empirical generalizations" related to customer lifetime value. We hope that other researchers — and practitioners — will confirm and extend some of our findings, and also uncover new behavioral patterns using the basic concepts and modeling platform presented here.

Appendix

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

The Pareto/NBD model is based on the following assumptions:

i. While active, the number of transactions made by a customer follows a Poisson process with transaction rate λ. This is equivalent to assuming that the time between transactions is distributed exponential with transaction rate λ.

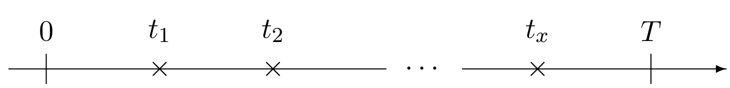
ii. Each customer has an unobserved "lifetime" of length τ (after which he is viewed as being "inactive"), which is exponentially distributed with dropout rate μ.

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

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

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

Let us assume we know 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:

i. 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 (inter-transaction-time) exponential density functions and the associated survival function:

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

ii. 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 information on when each of the x transactions occurred is not required; the only customer information required 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:

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

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

$$\begin{aligned} L(r, \alpha, s, \beta \,|\, X = x, t_x, T) &= \int_0^\infty \int_0^\infty L(\lambda, \mu \,|\, X = x, t_x, T) g(\lambda \,|\, r, \alpha) g(\mu \,|\, s, \beta) d\lambda d\mu \\ &= \frac{\Gamma(r+x)\alpha^r \beta^s}{\Gamma(r)} \left\{ \frac{1}{(\alpha+T)^{r+x}(\beta+T)^s} + \left(\frac{s}{r+s+x}\right) \mathsf{A}_0 \right\} \end{aligned} \qquad \text{(A1)}$$

where for $\alpha \geq \beta$

$$\begin{aligned} \mathsf{A}_0 = \frac{1}{(\alpha+t_x)^{r+s+x}} {}_2F_1\big(r+s+x, s+1; r+s+x+1; \tfrac{\alpha-\beta}{\alpha+t_x}\big) \\ - \frac{1}{(\alpha+T)^{r+s+x}} {}_2F_1\big(r+s+x, s+1; r+s+x+1; \tfrac{\alpha-\beta}{\alpha+T}\big) \end{aligned}$$

and for $\alpha \leq \beta$

$$\mathsf{A}_0 = \frac{1}{(\beta + t_x)^{r+s+x}} {}_2F_1\big(r+s+x, r+x; r+s+x+1; \tfrac{\beta-\alpha}{\beta+t_x}\big) - \frac{1}{(\beta + T)^{r+s+x}} {}_2F_1\big(r+s+x, r+x; r+s+x+1; \tfrac{\beta-\alpha}{\beta+T}\big)\,.$$

(See Fader and Hardie (2005) for details of the derivation.)

The four Pareto/NBD model parameters (r, α, s, β) can be estimated via 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

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

This can be maximized using standard numerical optimization routines; see Fader et al. (2005b).

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 becomes inactive (τ) occurs after T. Referring back to our derivation of the individual-level likelihood function, the application of Bayes' theorem gives us

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

We can now turn our attention to the derivation of DET (i.e., "discounted expected transactions") as implied by the Pareto/NBD model. The general explicit formula for the computation of customer lifetime value 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. Factoring 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 lifetime with continuous compounding at rate of interest δ is

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

(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) then 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 lifetime, (A3), times the probability that the customer with this purchase history is active at time T, (A2):

$$DET(\delta \,|\, \lambda, \mu, X = x, t_x, T) = DET(\delta \,|\, \lambda, \mu) P(\tau > T \,|\, \lambda, \mu, X = x, t_x, T)\,. \tag{A4}$$

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

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

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

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

Substituting (A2), (A4) and (A6) in (A5) gives us

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

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

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

where

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

and

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

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

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

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

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

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

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 (A9) and (A10) in (A8) gives us the following expression for DET as implied by the Pareto/NBD model:

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

[8] `http://functions.wolfram.com/07.33.02.0001.01`
[9] `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**, (1 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.

Berry, Michael J. A. and Gordon S. Linoff (2004), *Data Mining Techniques*, 2nd edition, Indianapolis, IN: Wiley Publishing, Inc.

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

Blattberg, Robert C., Gary Getz, and Jacquelyn S. Thomas (2001), *Customer Equity*, Boston, MA: 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), Part 2 of 2, S94–S107.

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

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

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2005b), "A Note on Implementing the Pareto/NBD Model in MATLAB," <`http://brucehardie.com/notes/008/`>.

Hughes, Arthur M. (2000), *Strategic Database Marketing*, 2nd edition, New York, NY: McGraw-Hill, Inc.

Malthouse, Edward C. (2003), "Scoring Models," in Dawn Iacobucci and Bobby Calder (eds.), *Kellogg on Integrated Marketing*, Hoboken, NJ: John Wiley & Sons, Inc., 227–249.

Malthouse Edward C. 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), Part 2 of 2, 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, NY: John Wiley & Sons, Inc.

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

Pfeifer, Phillip E., Mark E. Haskins, and Robert M. Conroy (2005), "Customer Lifetime Value, Customer Profitability, and the Treatment of Acquisition Spending," *Journal of Managerial Issues*, forthcoming.

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.

Reinartz, Werner and V. Kumar (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), 321–339.

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), 167–183.

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

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