More than meets the eye

BY PETER S. FADER, BRUCE G.S. HARDIE, AND KA LOK LEE

The move toward a customer-centric approach to marketing, coupled with the increasing availability of customer transaction data, has led to a heightened interest in the notion 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. However, the reality is more complex, as any analyst given the task of actually computing CLV will know. The key challenge is how to forecast a customer's future cash flows conditional on his or her past behavior.

Traditional Approaches

The standard approach taken by many analysts is to develop regression-type models that attempt to predict a customer's behavior as a function of his or her past behavior, as well as the veritable kitchen sink of other customer profile measures (e.g., demographics, mode of acquisition, length of relationship with the firm). The transaction database is split into two consecutive periods. Data from the second period are used to create the dependent variable for the model (e.g., buy/not-buy, number of transactions, total spend), whereas data from the first period are used to create the predictor variables. Period 1 behavior is frequently summarized in terms of each customer's "RFM" characteristics: recency (time of most recent purchase), frequency (number of past purchases), and monetary value (average purchase amount per transaction).

Executive Summary

Calculating customer lifetime value is complex, and the use of familiar regression-type models—which attempt to forecast future behavior based on only observable measures—is problematic and inadequate. A better approach is to perform the calculations using a probability model of buyer behavior, in which observed behavior is viewed as the outcome of a random process governed by latent characteristics. Companies that are serious about valuing their customer base must embrace this unconventional yet superior method.

But there are several problems with these models, especially when seeking to develop CLV estimates. First, the regression-type models are ad hoc; there is no well-established theory. Explanatory variables (including demographics, marketing variables, and behavioral measures such as RFM) are often added into the model simply because they lead to a higher R-squared. There is generally no compelling "story" behind many of these measures and their relationships with CLV. Is the fact that "they work" a good enough reason? A "curve-fitting" exercise might be an adequate way to try to explain past purchase patterns, but without a basis to justify the particular relationships uncovered by the model, it is hard to have faith in it to make predictions.

Second, regression models (and other forms of "data-mining" procedures) are designed to predict behavior in the next period. But when computing CLV, we are not only interested in period 2, but also need to predict behavior in periods 3, 4, 5, and so on. Having calibrated the regression model, we can predict period 3 behavior using the observed period 2 data. However, it is not clear how these models can be used to forecast buyer behavior for period 4 when we are unable to specify values for the RFM predictor variables in period 3, for each customer. As we move further away from the period for which we have actual values for the predictor variables, it becomes increasingly difficult to derive the expected value of the dependent variable. Long-term forecasts (which are required for a "lifetime" analysis) will be highly unreliable as a result.

Our third concern is more subtle, but at least as important as the other two. The developers of these models ignore the fact that the observed RFM variables are only imperfect indicators of underlying behavioral characteristics; they are not fixed variables such as demographics. Different "slices" of the data will yield different values of the RFM variables, and therefore different forecasts. Small changes in past behavior could lead to dramatic differences in the model's expectation of future valuation. We need to explicitly account for the stochastic nature of these measures, but the standard regression framework does not allow us to do so.

Despite these concerns, regression-like models continue to comprise the majority of all CLV modeling efforts. In contrast, we now discuss another general class of methods that can be used for CLV purposes, while overcoming all of the previously mentioned problems (and therefore giving the manager more faith in its validity and applicability).

An Alternative Approach

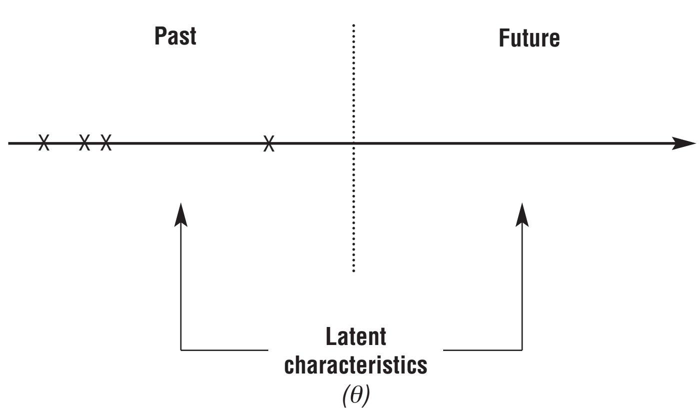
A probability modeler approaches the problem with the mind-set that observed behavior is the outcome of an underlying stochastic process. That is, we have only a "foggy window" as we attempt to see our customers' true behavioral tendencies, and therefore the past is not a perfect mirror of the future. For instance, if a customer made two purchases last year, then is he or she necessarily a "two per year" buyer, or is there some chance that he or she might make three or four or perhaps even zero purchases next year? With this kind of uncertainty in mind, we wish to focus more on the latent process that drives these observable numbers, rather than the observables themselves.

As illustrated in Exhibit 1, the transactions associated with a customer—those observed in the past and those from the yet-to-be-observed future—are a function of that customer's underlying behavioral characteristics (denoted by θ).

By specifying a mathematical model in which the observed behavior is a function of an individual's behavioral characteristics (i.e., past = $f(\theta)$), an application of Bayes' theorem enables us to make inferences about an individual's latent characteristics ($\hat{\theta}$) given his or her observed behavior. We can then make predictions regarding behavior as a function of the inferred latent characteristics.

In contrasting this two-step approach ($\hat{\theta}$ = f(past) and future = $f(\hat{\theta})$) with the single-step regression model (future = f(past)), we find that the use of a formal probability model avoids all the shortcomings associated with regression-type models. First, there is no need to split the observed transaction data into two periods to create a dependent variable; we can use all of the data to make inferences about the customer's behavioral characteristics. Second, we can predict behavior over future time periods of any length; we can even derive an

Exhibit 1 A probability modeler's view of transactions



explicit expression for CLV over an infinite horizon (with discounting to acknowledge the lower present value of purchases that occur in the distant future). And third, the recognition that the observed behavior—and therefore the observed RFM variables—is only an imperfect indicator of underlying behavioral characteristics gives us more confidence that we are capturing the actual process that is generating the customer's past, present, and future behavior.

A Specific Model

To illustrate these ideas, we briefly describe a model that has been developed to compute CLV in a continuous-time noncontractual setting. This model is developed more fully in our November 2005 article in the *Journal of Marketing Research*, "RFM and CLV: Using Iso-value Curves for Customer Base Analysis" (hereafter "FHL" for our initials), which conveys all of the required technical details as well as a more extensive empirical analysis than we provide here. Our objective in this article is to bring across its essential elements and managerial intuition. At the heart of any modeling effort is a "story" regarding buyer behavior. So let's consider the following set of assumptions that comprises the basis of our story.

We first assume that the amount spent per transaction is independent of the transaction process. This means our model of buyer behavior can be separated into a sub-model for the flow of transactions and a sub-model for revenue per transaction.

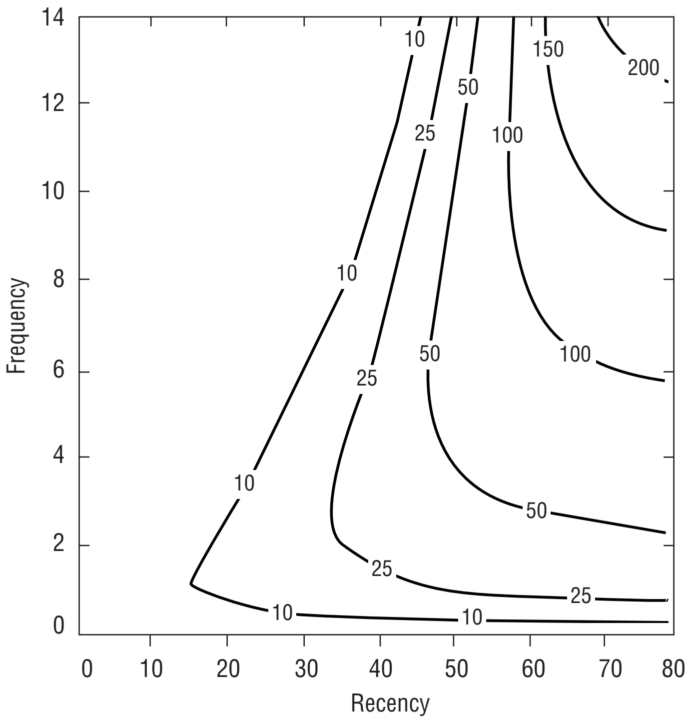
Our model for the transaction stream is based on the following assumptions:

- A customer's relationship with the firm has two phases: He or she is "alive" for an unobserved period of time, and then becomes permanently inactive. But this time of inactivity need not be a short- or medium-term occurrence; in some cases, it might not arise at all during the customer's physical lifetime.
- While alive, a customer "randomly" purchases around his or her mean transaction rate.
- Both the transaction rates and dropout rates vary across customers.

Our model for the spend process is based on the following assumptions:

- The dollar value of a customer's given transaction randomly varies around his or her mean transaction value.
- Mean transaction values vary across customers but do not vary over time for any given individual.

Exhibit 2 Iso-value plot



In the formal model derivation presented in FHL, we translate this verbal model of buyer behavior into the language of math, and show a critically important result: We do not need to know a customer's entire transaction history; RFM are sufficient summary measures for all of his or her past behavior! This means we are able to make inferences about an individual's latent characteristics (i.e., transaction rate, dropout rate, mean transaction value) based on only his or her RFM inputs.

Recall that CLV is defined as the present value of the future cash flows associated with a customer. A consequence of our assumption that monetary value is independent of the underlying transaction process is that the net cash flow per transaction can be factored out of the calculation, which means we focus on forecasting the "flow" of transactions (discounted to yield a present value). This number of discounted expected transactions (DET) can then be rescaled by a net cash flow "multiplier" to yield an overall estimate of expected CLV: E(CLV) = E(net cash flow/transaction) × DET.

This decomposition offers two significant benefits. First, it breaks down and simplifies the computational steps associated with the model; it is easier to perform all the required calculations for each component separately and then combine, instead of trying to disentangle all of the factors at one time. Second, it offers diagnostic benefits that can assist a firm in identifying problem areas and determining how to allocate marketing resources to address them.

Given the mathematical expressions that characterize the duration of a customer's "lifetime" and transaction behavior while alive, we are able to derive an expression for DET as a function of an individual's (unobserved) transaction and dropout rates. Similarly, we can derive an expression for the expected net cash flow/transaction as a function of an individual's (unobserved) mean transaction value. As we can make inferences about an individual's latent behavioral characteristics given his or her RFM, we can derive specific formulas of the form DET = f_1(R, F, transaction model parameters) and E(net cash flow/transaction) = f_2(F, M, spend model parameters)—where the transaction model parameters describe how the transaction and dropout rates vary across customers, and the spend model parameters describe how the mean transaction values vary across customers. This gives us an expression of the form E(CLV) = f_3(R, F, M, model parameters).

At first glance, this kind of expression might seem to be no different from what might emerge from a regression-like model. But by building up to this expression from our basic individual-level story, we overcome the concerns described in the "Traditional Approaches" section. Furthermore, this expression is very flexible and can capture many different kinds of relationships (including some fairly counterintuitive ones) among the set of RFM variables. We now turn to a specific application that will reveal some of these patterns.

An Illustrative Application

We draw our data set from the online music site, CDNOW. We consider the cohort of 23,570 individuals who made their first-ever purchase at the site in the first quarter of 1997. We have data covering their initial and repeat purchases up to the end of June 1998. Having estimated the transaction and spend model parameters using all 78 weeks of data, we can compute DET for each individual as a function of their recency and frequency. When we also bring in our sub-model for spending per transaction, we have a complete picture of CLV for the entire customer base. In FHL we demonstrate some model diagnostics and validity tests for each of the sub-models, but here we just jump right to the CLV analysis.

Exhibit 2 presents an "iso-value" plot of expected CLV as a function of recency (the week of last purchase, ranging from 1 to 78) and frequency (the number of repeat purchases made in the 78-week period, ranging from 0 to 14)—assuming a 15% annual discount rate and 30% margin, and a historic average transaction value of \$20. Each curve in the graph links customers with equivalent future value, despite differences in their past behaviors.

For the high-value customers (i.e., upper right of Exhibit 2), it is clear that recency and frequency each have a direct and positive association with CLV. 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 who made seven repeat purchases with the last one occurring in week 35 (frequency = 7, recency = 35) has an approximate CLV of \$10, the same as someone with a single repeat purchase that occurred in week 20 (frequency = 1, recency = 20). In general, for people with low recency, higher frequency seems to be a bad thing. Initially, this might seem like a mistake in the model, but upon further reflection, it starts to make sense. If we knew for sure that both customers were still active in week 78, then we would expect the customer who made seven repeat purchases to have a greater CLV, in light of his or her higher number of past purchases. However, his or her RFM profile (specifically, high frequency but low recency) suggests that—most likely—he or she is no longer active at the end of week 78. On the other hand, the second customer has a lower underlying purchase rate, so it is reasonably likely that he or she is still active by week 78, even though he or she hasn't made a purchase for the past 58 weeks. The net effect is that both customers are estimated to have the same CLV, despite their very different past purchase histories.

The overall clarity provided by this exhibit demonstrates the usefulness of a formal model for understanding the relationship between CLV and RFM. Furthermore, the backward-bending iso-value curves emphasize the importance of using a model with sound behavioral assumptions, instead of an ad hoc regression-type model that would probably miss this pattern and lead to faulty inferences for a large portion of the recency-frequency space.

Another way of understanding and appreciating the usefulness of our CLV estimation method is to combine the model-driven RFM-CLV relationship (as presented in the iso-value curves) with the actual RFM patterns seen in our data set, to get a sense of the overall value of this cohort of customers. To enhance the clarity and interpretability of this combination, we will segment the customers on the basis of their RFM characteristics and report the average expected CLV of each segment. (This allows us to "close the loop" with traditional RFM segmentation analyses, and show how our model can be employed for target marketing purposes.)

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

We compute each customer's expected CLV (conditional on his or her past behavior) and compute segment-level averages. In Exhibit 3 we show the average expected CLV for each of the 28 segments; the size of each RFM segment is reported in parentheses. Multiplying the average E(CLV) by the size of the segment gives us the expected residual value of the group of customers in that segment. 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 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 applied to such a large group of customers. Many managers would assume that after a year and a half of inactivity, a customer has dropped out of his or her 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.

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 by 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—is the most valuable, with a total expected lifetime value of $414,938 beyond week 78. This represents nearly 38% of the future value of the entire cohort (which is worth just more than $1.1 million).

Exhibit 3 Average E(CLV) by RFM group

		Recency			
	Frequency	0	1	2	3
M=0	0	$4.40 (12,054)			
M=1	1		$6.39 (1,197)	$20.52 (482)	$25.26 (71)
	2		$7.30 (382)	$31.27 (488)	$41.55 (419)
	3		$4.54 (57)	$48.74 (256)	$109.32 (484)
M=2	1		$9.02 (650)	$28.90 (264)	$34.43 (68)
	2		$9.92 (358)	$48.67 (545)	$62.21 (414)
	3		$5.23 (86)	$77.85 (478)	$208.85 (972)
M=3	1		$16.65 (676)	$53.20 (371)	$65.58 (57)
	2		$22.15 (329)	$91.09 (504)	$120.97 (396)
	3		$10.28 (101)	$140.26 (447)	$434.95 (954)

Exhibit 4 Classifying customer bases

Opportunities for transactions	Noncontractual	Contractual
Continuous	Grocery purchases Doctor visits Hotel stays	Credit card Student meal plan Cell phone usage
Discrete	Event attendance Prescription refills Charity fund drives	Magazine subscriptions Insurance policy Health club memberships

Type of relationship with customers

Other Types of Customer Bases

We have argued that the task of forecasting a customer's future cash flows conditional on his or her past behavior is best tackled using probability models, rather than traditional regression-type models. To illustrate the logic of such a probability modeling approach, we briefly described a model developed to compute CLV in a continuous-time noncontractual setting.

Many readers will have glanced over the words "continuous-time" and "noncontractual" without reflecting on their significance. We see this as a common problem in customer-base analysis (and in the use and teaching of CLV in particular). Managers and educators often rely on a "one size fits all" mind-set, rather than thinking carefully about the different kinds of relationships that different kinds of firms have with their customers.

In stating "continuous-time," we mean that transactions can occur at any point in time. This is in contrast to discrete-time, where the transactions can occur only at a discrete point in time. For example, business relationships ranging from special event attendance to prescription refills and contract renewals are often restricted to take place during specific periods of time. In these situations, the key managerial question isn't when the transactions will take place (as in the case of the CDNOW data set), but whether a transaction will take place during the specified period.

The word "noncontractual" refers to a different aspect in the relationship between the firm and its customers, and obviously contrasts with a contractual setting. In the latter case, the time when the customer becomes inactive is observed (e.g., when the customer fails to renew his or her subscription, or contacts the firm to cancel his or her contract). In the former case (which includes the CDNOW example), the point when the customer becomes inactive is not observed by the firm. This means that the firm cannot differentiate between a customer who has ended his or her relationship with the firm and one who is merely in the midst of a long hiatus between transactions. In noncontractual settings, we cannot meaningfully utilize notions such as "retention rates" or apply tools such as survival analysis.

These two dimensions lead to a classification of customer bases as illustrated in Exhibit 4. The probability model we have outlined was developed to compute CLV for those business contexts that can be placed in the top-left quadrant. Although different models are required to compute CLV for each of the other three quadrants, they are easy to construct using the tools of probability modeling. The resulting formulas will vary, but they will all share the basic logic that went into the model presented here. We have research projects under way to lay out the mathematical details (and unique managerial challenges) associated with these other three quadrants.

Embrace the Unfamiliar

Does the framework described earlier offer a complete picture of how to model CLV? Not really. We haven't mentioned a number of considerations such as customer acquisition, the incorporation of covariates (demographics and marketing mix effects), and capturing differences across multiple cohorts of customers. Although these are important issues, they can all be handled using the probability modeler's tool kit. In fact, once someone understands the value of focusing on the latent characteristics instead of the observable measures, it becomes quite natural to see how these additional components can be brought into the framework. For instance, instead of assuming that marketing mix effects directly influence a customer's transaction pattern (as in a regression model), we suggest that these effects influence his or her latent characteristics (e.g., purchase propensity and/or dropout rate)—which in turn change the probability that a purchase takes place. This helps to give us a cleaner read on the incremental impact of, say, a price change or the introduction of a new loyalty program.

Thus, it isn't our goal to draw the complete picture here, but simply to help the sophisticated manager appreciate this somewhat unconventional (but powerful) way to think about calculating CLV. Companies that are serious about valuing their customer base have too much at stake to rely on models that are familiar, but inadequate for the task at hand. We hope that our observations will encourage them to approach customer-base analysis exercises from this promising new direction. ●

Peter S. Fader is the Frances and Pei-Yuan Chia Professor of Marketing at the Wharton School of the University of Pennsylvania in Philadelphia, and may be reached at faderp@wharton.upenn.edu. **Bruce G.S. Hardie** is associate professor of marketing at London Business School, and may be reached at bhardie@london.edu. **Ka Lok Lee** is market research analyst at Catalina Health Resource in Blue Bell, Penn., and may be reached at kaloklee@alumni.upenn.edu.