

Customer Lifetime Value Measurement

Sharad Borle, Siddharth S. Singh

Jesse H. Jones Graduate School of Management, Rice University, Houston, Texas 77005
{sborle@rice.edu, sssingh@rice.edu}

Dipak C. Jain

J. L. Kellogg School of Management, Northwestern University, Evanston, Illinois 60208,
d-jain@kellogg.northwestern.edu

The measurement of customer lifetime value is important because it is used as a metric in evaluating decisions in the context of customer relationship management. For a firm, it is important to form some expectations as to the lifetime value of each customer at the time a customer starts doing business with the firm, and at each purchase by the customer. In this paper, we use a hierarchical Bayes approach to estimate the lifetime value of each customer at each purchase occasion by jointly modeling the purchase timing, purchase amount, and risk of defection from the firm for each customer. The data come from a membership-based direct marketing company where the times of each customer joining the membership and terminating it are known once these events happen. In addition, there is an uncertain relationship between customer lifetime and purchase behavior. Therefore, longer customer lifetime does not necessarily imply higher customer lifetime value.

We compare the performance of our model with other models on a separate validation data set. The models compared are the extended NBD–Pareto model, the recency, frequency, and monetary value model, two models nested in our proposed model, and a heuristic model that takes the average customer lifetime, the average interpurchase time, and the average dollar purchase amount observed in our estimation sample and uses them to predict the present value of future customer revenues at each purchase occasion in our hold-out sample. The results show that our model performs better than all the other models compared both at predicting customer lifetime value and in targeting valuable customers. The results also show that longer interpurchase times are associated with larger purchase amounts and a greater risk of leaving the firm. Both male and female customers seem to have similar interpurchase time intervals and risk of leaving; however, female customers spend less compared with male customers.

Key words: customer lifetime value; customer equity; hierarchical Bayes

History: Accepted by Jagmohan S. Raju, marketing; received November 19, 2004. This paper was with the authors 11 months for 3 revisions. Published online in *Articles in Advance* December 11, 2007.

1. Introduction

The focus of firms on customer relationship management (CRM) in recent years to achieve higher profitability has resulted in the popularity of various firm initiatives to retain customers and increase purchases by them (Jain and Singh 2002, Dowling and Uncles 1997, O'Brien and Jones 1995). In the context of customer relationship management, customer lifetime value (CLV), or customer equity, becomes important because it is a metric to evaluate marketing decisions (Blattberg and Deighton 1996).

For a firm, it is of interest to know how much net benefit it can expect from a customer *today*. Therefore, at each point in a customer's lifetime with the firm, the firm would like to form some expectation regarding the lifetime value of that customer. This expectation can then be used to make marketing activities more efficient and effective. In light of the fact that marketing budgets are limited, a firm's strategy of focusing different types of marketing instruments

on different customers based on their expected value can help the firm get better return on its marketing investment.

To do this, a critical problem faced by a firm is the measurement of the CLV. Researchers have suggested various methods to use customer-level data to measure the CLV (Fader et al. 2005, Rust et al. 2004, Berger and Nasr 1998, Schmittlein and Peterson 1994). In measuring customer lifetime value, a common approach is to estimate the present value of the net benefit to the firm from the customer (generally measured as the revenues from the customer minus the cost to the firm for maintaining the relationship with the customer) over time (Blattberg and Deighton 1996). Typically, the cost to the firm for maintaining a relationship with its customers is controlled by the firm, and therefore is more predictable than the other drivers of CLV. As a result, researchers generally consider a customer's revenue stream as the benefit from the customer to the firm.

It is noteworthy that research on CLV measurement has so far focused on specific contexts. This is necessary because the data available to a researcher or firm in different contexts might be different. The two types of context generally considered are noncontractual and contractual (e.g., Reinartz and Kumar 2000, 2003). A *noncontractual* context is one in which the firm does not observe customer defection, and the relationship between customer purchase behavior and customer lifetime is not certain (e.g., Fader et al. 2005; Schmittlein and Peterson 1994; Reinartz and Kumar 2000, 2003). A *contractual* context, on the other hand, is one in which customer defections are observed, and longer customer lifetime implies higher customer lifetime value (e.g., Thomas 2001, Bolton 1998, Bhattacharya 1998). The context of our study, as we describe later, has elements of both contractual and noncontractual settings, a scenario that has not been analyzed in-depth previously (Singh and Jain 2007).

Different models for measuring CLV arrive differently at estimates of the expectations of future customer purchase behavior. For example, some models consider discrete time intervals and assume that each customer spends a given amount (e.g., an average amount of spending in the data) during each interval of time. This information, along with some assumption about the customer lifetime length, is used to estimate the lifetime value of each customer by a discounted cash-flow method (Berger and Nasr 1998). In another model, Rust et al. (2004) combine the frequency of category purchases, average quantity of purchase, brand-switching patterns, and the firm's contribution margin to estimate the lifetime value of each customer. Because customer purchase behavior might change over a customer's lifetime with the firm, methods that incorporate past customer behavior to form an expectation of future customer behavior and, subsequently, the remaining customer lifetime value are likely to have advantages over other methods (e.g., Schmittlein and Peterson 1994).

A popular method that follows such an approach in a noncontractual context is the negative binomial distribution (NBD)–Pareto model by Schmittlein et al. (1987). In this model, past customer purchase behavior is used to predict the future probability of a customer remaining in business with the firm (the probability of each customer being *alive*). Along with a measure of purchase frequency and amount spent during a purchase, this probability can be used to estimate customer lifetime value (Reinartz and Kumar 2000, 2003; Schmittlein and Peterson 1994). The NBD–Pareto model is applied in instances where customer lifetimes are not known with certainty, i.e., it is not known when a customer stops doing business with a firm; the model assumes that individual customer

lifetimes with the firm are exponentially distributed. As discussed by Schmittlein and Peterson (1994), in contexts (such as ours) where customer lifetimes are observed, the NBD–Pareto model has limitations and is not suitable.

Another approach that can naturally incorporate past behavioral outcomes into future expectations is a Bayesian approach (Rossi and Allenby 2003). Bayesian methods can incorporate such prior information in the structure of the model easily through the priors of the distributions of the drivers of CLV. Furthermore, this approach can be used in any context. Therefore, we use such an approach to measure customer lifetime value, leveraging the extra information available to the firm in observing customer lifetimes. A hierarchical Bayesian model is developed that jointly predicts a customer's risk of defection and spending pattern at each purchase occasion. This information is then used to estimate the lifetime value of each customer of the firm at every purchase occasion. We compare the predictions from our model on a separate validation sample to those obtained from some extant methods of measuring CLV, namely, the extended NBD–Pareto framework,¹ a heuristic method, and two models nested in our proposed model. We also compare the performance of our model in targeting customers with the performance of a recency, frequency, and monetary value (RFM) framework, in addition to the other models mentioned previously.

The results show that our proposed model performs better in terms of predicting customer lifetime value and also in targeting valuable customers than the methods used for comparison. We find that customers' purchase timing, purchase amount, and risk of defecting are not independent of each other, which validates our joint modeling approach.

The remainder of this paper is organized as follows: the next section describes the data, §3 details the model development, §4 discusses the estimates, and §5 applies the model to a separate validation sample data set and compares its performance with other methods. Finally, §6 ends the paper with a summary and discussion of the results.

2. The Data

The data come from a membership-based direct marketing company. Examples of such companies are membership-based clubs such as music clubs, book clubs, and other types of purchase-related clubs. The membership is open to the general public.² Information about any purchase by a customer is known to

¹ Proposed by Schmittlein et al. (1987) and later extended by Schmittlein and Peterson (1994).

² Due to a data confidentiality agreement with the company, we are unable to divulge more details about the company.

the firm only when the purchase happens. Similarly, customer lifetime length (total membership duration) with the firm is not known to the firm until a customer leaves the firm (i.e., the customer terminates her membership). In such firms, both the purchase timing and spending on purchases do not happen continuously or at known periods, and can only be predicted probabilistically. Therefore, the data most closely resemble a noncontractual context except that customer lifetime information of past customers is known to the firm with certainty (i.e., the time when a membership begins and the time when it ends are known once these events happen for each customer).

The data consist of two random samples, both drawn (without replacement) from the population of all the customers who joined the firm in a specific year in the late 1990s. They contain information about all the purchases by customers from the date of the start of their membership, i.e., joining the firm, until the termination of their membership.³

The first part of the data, referred to as the *estimation sample*, contains 1,000 past customers and consists of a total of 7,108 purchase occasions. It traces the purchase behavior of these customers over their entire lifetime with the firm. The dates of membership initiation and termination are known for each customer, i.e., completed lifetime lengths are known for each customer in the data. The second part, consisting of another 500 past customers (a *validation sample*), was selected for predictive testing and to illustrate the application of the model. The data contain three dependent measures of primary interest viz. the interpurchase times (*TIME*), the purchase amounts (*AMNT*), and the customer lifetime information (total membership duration of each customer). Figures 1 and 2 display histogram plots of the interpurchase times and purchase amounts, respectively, across all purchase occasions for the estimation sample.

On average, a customer takes about 9 to 10 weeks between purchases. The bulk of the purchases (more than 90%) occur within 20 weeks of the previous purchase. However, as much as 2% of all purchases occur with interpurchase times in excess of 35 weeks. In terms of purchase amounts, again there is considerable heterogeneity in the population. On average, a purchase costs about \$17, with the bulk of purchases (more than 90% of all purchases) being less than \$30. However, we do observe about 2% of all purchases to be in excess of \$50.⁴

³ Note that we do not have any censored observation of customer lifetime. This is because by the time we received the data, all of the customers in the entire relevant population (from which the samples were drawn) had terminated their memberships.

⁴ We use the dollar (\$) as a general unit of currency.

Figure 1 Interpurchase Times

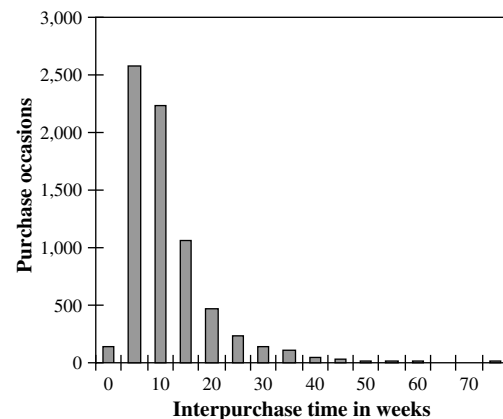


Figure 2 Purchase Amounts

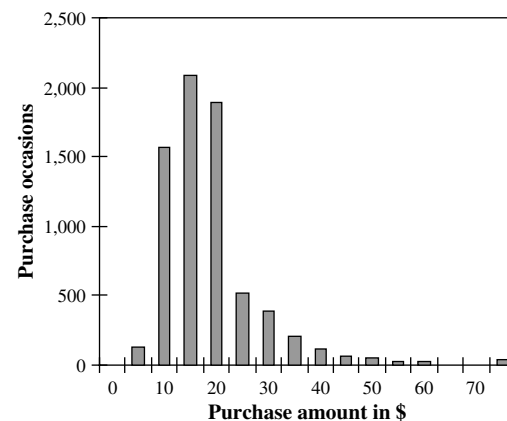


Table 1 presents summary statistics of the variables used in the estimation sample.

The other variables we use are a dummy variable, *GENDER*, representing the gender of a customer (female = 1; there are 67% female customers in the sample) and the lag values of interpurchase times and purchase amounts.

Figure 3 below is a histogram of the lifetimes observed across customers in our estimation sample, and Table 2 contains some corresponding summary statistics.

The lifetime plot (Figure 3) shows significant heterogeneity across customers. The customer lifetime varies from less than 10 weeks to over 240 weeks, the average being about 82 weeks. The firm also observes the “exit pattern” of customers, i.e., which customer

Table 1 Summary Statistics

	<i>TIME</i> (weeks)	<i>AMNT</i> (\$)	<i>GENDER</i> (0: Male)
Mean	9.43	16.98	0.67
Std. dev.	8.90	10.76	0.47
Minimum	0	0.50	0
Maximum	128	265.86	1

Figure 3 Customer Lifetimes

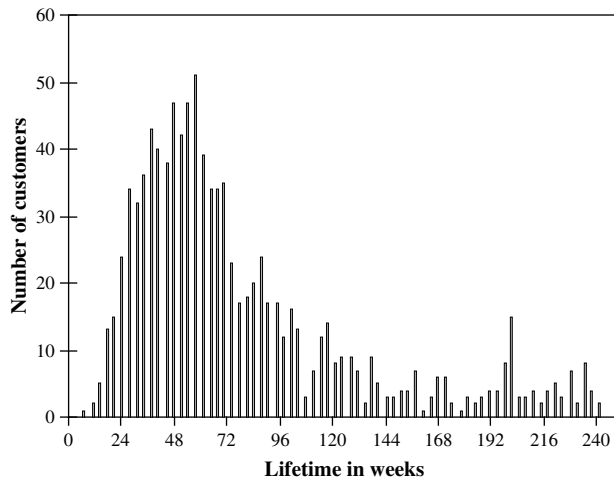


Table 2 Some Summary Statistics on Lifetime Distribution

	LIFETIME (weeks)
Mean	82.0
Std. dev.	54.8
Minimum	7
Maximum	251

left after making the first purchase, the second purchase, the third purchase, and so on. Figures 4(a) and 4(b) display the histogram plot and the corresponding hazard of this exit pattern of customers, respectively. This is the third dependent quantity of interest and captures the customer mortality information. The horizontal axis in both figures is the number of purchase occasions; in our estimation sample, we observe a maximum of 41 purchase occasions.⁵ The vertical axis in Figure 4(a) is the number of customers who terminate their membership with the firm after a particular purchase occasion. The vertical axis in Figure 4(b) is the average probability of a customer defecting (the hazard rate) given that the customer has survived until a particular purchase occasion.

Figure 4(b) also contains a third-degree polynomial approximation of the actual hazard pattern (the dotted line). An interesting facet about the empirical hazard pattern in Figure 4(b) is that the hazard rises until the sixth purchase occasion and then decreases until about the 17th purchase occasion and subsequently rises again. It is conceivable that people join the firm, try it out for a few occasions, and then some of the customers decide to quit the firm whereas others become consistent purchasers.

⁵ The maximum number of times any customer bought from the firm was 41.

Figure 4(a) Number of Customers Existing After a Particular Purchase

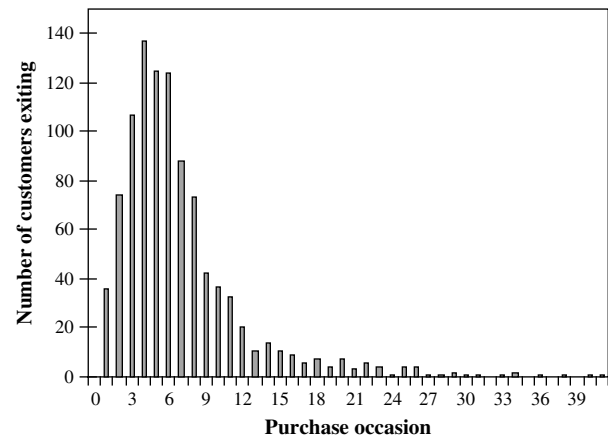
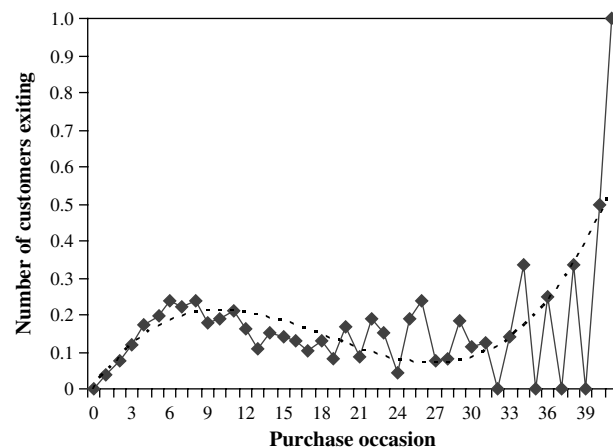


Figure 4(b) The Corresponding Hazard Pattern



In the next section we introduce our model and subsequently apply it to predict the customer lifetime values at each purchase occasion.

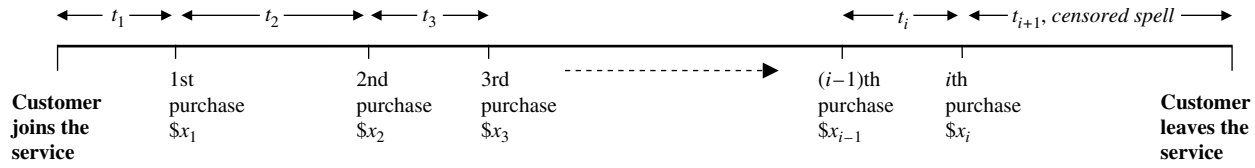
3. The Model

Typical data for each customer can be depicted as in Figure 5. A customer joins the firm, makes her first purchase of $\$x_1$ after t_1 weeks, makes her second purchase of $\$x_2$ after another t_2 weeks, and so on until the i th purchase occasion. Subsequently, the customer leaves with a censored spell of t_{i+1} weeks.

We develop a joint model of the three dependent quantities of interest viz. the *interpurchase time*, the *purchase amount*, and the *probability of leaving* given that a customer has survived a particular purchase occasion (i.e., the hazard rate⁶ or the risk of defection). We specify models for interpurchase time, purchase amounts, and the risk of defection and then allow a correlation structure across these three models, thus leading to a joint model of these three quantities. The

⁶ See Jain and Vilcassim (1991) for an exposition of hazard models.

Figure 5 Visual Depiction of a Typical Data String



model is then jointly estimated and we use the estimates to predict the customer lifetime value at each purchase occasion for every customer in the validation sample.

3.1. Interpurchase Time Model

The interpurchase time is measured in weeks and we assume that it follows an NBD process, i.e.,

$$TIME_{hi} \sim NBD(\lambda_{hi}, \nu_1), \quad (1)$$

where $TIME_{hi} = 0, 1, 2, 3, \dots$ measures the interpurchase time in weeks for customer h at purchase occasion i (the time between the $(i-1)$ th and the i th purchase occasion), and (λ_{hi}, ν_1) are the parameters of the NBD distribution. The parameter λ_{hi} is the mean of the distribution and ν_1 is the dispersion parameter. The NBD is a well known and used distribution in the marketing literature. It is a generalization of the Poisson distribution and is useful in modeling over-dispersed count data. Another flexible distribution to model over-dispersed data is the COM-Poisson distribution (Boatwright et al. 2003); however, in our application the NBD outperformed the COM-Poisson in its predictive ability.⁷

The probability mass function of the NBD distribution is as follows:

$$P(TIME_{hi}/\lambda_{hi}, \nu_1) = \frac{\Gamma(\nu_1 + TIME_{hi})}{\Gamma(\nu_1)\Gamma(TIME_{hi} + 1)} \cdot \left(\frac{\nu_1}{\nu_1 + \lambda_{hi}}\right)^{\nu_1} \left(\frac{\lambda_{hi}}{\nu_1 + \lambda_{hi}}\right)^{TIME_{hi}}. \quad (2)$$

Thus, the likelihood contribution of a complete spell is as given in Equation (2) whereas the likelihood contribution of a censored spell is as follows:

$$1 - \sum_{r=0}^{TIME_{hi}} P(r/\lambda_{hi}, \nu_1). \quad (3)$$

We further specify the parameter λ_{hi} as follows:

$$\log \lambda_{hi} = \lambda_h + \lambda_i + \lambda_{1h} \log \log TIME_{hi} + \lambda_2 GENDER_h$$

$$\text{where } \lambda_i = \lambda'/i + \lambda''/i^2. \quad (4)$$

The variable $GENDER_h$ is the gender of customer h (female = 1; male = 0). The coefficient on $GENDER_h$ addresses any gender differences in the population in terms of purchase frequencies (interpurchase times). The quadratic trend parameter $\lambda_i (= \lambda'/i + \lambda''/i^2)$ allows for nonstationarity in the interpurchase times across purchase occasions.⁸

The parameter λ_{1h} specifies the impact of lag interpurchase time⁹ on the current interpurchase time. We incorporate heterogeneity over this parameter by specifying a normal distribution for the λ_{1h} values:

$$\lambda_{1h} \sim Normal(\bar{\lambda}_1, \tau_1^2). \quad (5)$$

3.2. Purchase-Amount Model

The amount (in dollars, used as a general unit of currency) expended by customer h on purchase occasion i is denoted by $AMNT_{hi}$. We assume that this variable follows a log-normal process. Thus, we have

$$\log AMNT_{hi} \sim Normal(\mu_{hi}, \sigma^2), \quad (6)$$

where (μ_{hi}, σ^2) are the parameters (mean and variance, respectively) of the distribution. An analogous structure (analogous to the interpurchase-time model, Equations (4) and (5)) is allowed for the μ_{hi} parameter as follows:

$$\mu_{hi} = \mu_h + \mu_i + \mu_{1h} \log \log AMNT_{hi} + \mu_2 GENDER_h,$$

$$\text{where } \mu_i = \mu'/i + \mu''/i^2. \quad (7)$$

The coefficient μ_2 specifies the impact of gender on purchase amounts and the coefficient μ_i allows for a nonlinear trend in the purchase amounts across purchase occasions. The coefficient μ_{1h} specifies the impact of lagged dollars spent on future amounts expended. We allow this parameter to vary across customers as follows:

$$\mu_{1h} \sim Normal(\bar{\mu}_1, \tau_2^2). \quad (8)$$

⁷ Although, we must point out that NBD may not dominate over the COM-Poisson in all applications. Where under-dispersion is prevalent in the data, the COM-Poisson will dominate over the NBD (Borle et al. 2007). Even in over-dispersed data, in some applications the COM-Poisson would give a better fit (see Shmueli et al. 2005).

⁸ Here, i indexes the purchase occasion. Higher-order polynomials (beyond quadratic) were also estimated and not found to be statistically significant.

⁹ In a few instances (less than 0.5% of the data) where the lag interpurchase time is 0, we replace the value with 1.

3.3. Customer-Defection Model

The hazard of lifetime $h(LIFE_{hi})$ for customer h is the risk of leaving in the i th spell (probability that the customer after having made the $(i - 1)$ th purchase will leave the firm without making the i th purchase). We use a discrete-hazard approach to model this probability (see Singer and Willett 2003):

$$h(LIFE_{hi}) = \{1 + \exp(-\delta_{hi})\}^{-1}. \quad (9)$$

Retaining the general structure of the earlier two models (interpurchase-time and purchase-amount models), we specify δ_{hi} in Equation (9) as follows:

$$\begin{aligned} \delta_{hi} = & \delta_h + \delta_i + \delta_{1h} \log \log TIME_{hi} \\ & + \delta_{2h} \log \log AMNT_{hi} + \delta_3 GENDER_h, \\ \text{where } \delta_i = & \delta' i + \delta'' i^2 + \delta''' i^3. \end{aligned} \quad (10)$$

Nonstationarity across purchase occasions is incorporated in the discrete-hazard function by a third-order polynomial expression, $\delta_i = \delta' i + \delta'' i^2 + \delta''' i^3$, where i indexes the purchase occasion. Such a third-degree polynomial expansion is a parsimonious yet useful alternative to specifying coefficients for each purchase occasion in the discrete hazard. We observe a total of 41 purchase occasions in our data, so one alternative could have been to specify 41 separate coefficients for each purchase occasion. This would, however, hinder prediction beyond 41 purchase occasions. Therefore, we use three coefficients to specify a polynomial time trend.¹⁰ The other variables in the equation are the lagged interpurchase times, the lagged purchase amounts, and the gender variable.¹¹ We specify a heterogeneity structure over the coefficients for the lagged variables as follows:

$$\delta_{1h} \sim Normal(\bar{\delta}_1, \tau_3^2); \quad (11)$$

$$\delta_{2h} \sim Normal(\bar{\delta}_2, \tau_4^2). \quad (12)$$

The intercept δ_h in Equation (10) can be interpreted as a measure of the baseline risk of defection for customer h ; this risk is then further modified by the time trend (the polynomial expression) and the other covariates. The pattern of these estimates is indicative of the risk of defection in the population at various purchase occasions and is helpful to the firm in its targeted marketing activities.

¹⁰ Higher-order polynomial terms (beyond third order) were not found to be statistically significant.

¹¹ Interaction of the gender variable with the lagged variables was also explored in all three of the models. None of the interactions were found to be “significant.”

3.4. A Correlation Structure

To allow the three dependent variables (interpurchase time, purchase amount, and the risk of defection) to be related to each other, we allow a correlation structure across the three models specified in §§3.1–3.3. The correlations across the three equations (Equations (4), (7), and (10)) are introduced as follows:

$$\mathbf{\Omega}_h \sim MVNormal(\bar{\mathbf{\Omega}}, \mathbf{\Sigma}), \quad (13)$$

where $\mathbf{\Omega}_h = [\lambda_h, \mu_h, \delta_h]'$; the parameters $\lambda_h, \mu_h, \delta_h$ are as specified in Equations (4), (7), and (10), respectively. Furthermore, $\bar{\mathbf{\Omega}} = [\bar{\lambda}, \bar{\mu}, \bar{\delta}]'$ and $\mathbf{\Sigma}$ is a 3×3 variance-covariance matrix. The off diagonal elements of the $\mathbf{\Sigma}$ matrix specify the structure of covariance across the three variables in the respective models (i.e., interpurchase time, purchase amount, and the risk of defection). Incorporating such a covariance structure allows for dependencies across the three outcomes and is an efficient use of information in the data.

3.5. Estimation

There are three models to be jointly estimated: the interpurchase time, the purchase amount, and the customer-defection model (Equations (1)–(13)). The Bayesian specification across the three models is completed by assigning appropriate prior distributions on the parameters to be estimated. The models are estimated using a Markov Chain Monte Carlo (MCMC) sampling algorithm. The details of the prior distributions used in the analysis and the estimation algorithm can be obtained from the authors.

4. The Estimated Coefficients

The estimation result is a posterior distribution for each of the parameters. These are summarized by their posterior means and standard deviations. Tables 3(a), 3(b), and 3(c) report these estimates for parameters that are not specific to individual customers (for the interpurchase time, the purchase amount, and the customer-defection models, respectively). Furthermore, Table 4 reports the estimated covariance structure across the three models. The figures in parentheses are the posterior standard deviation and the superscript asterisks indicate that the 95% posterior interval for the parameter does not contain 0. This is interpreted as an indicator of the estimate being statistically different from zero.

The parameters $\bar{\mathbf{\Omega}} = [\bar{\lambda}, \bar{\mu}, \bar{\delta}]'$ and the 3×3 variance-covariance matrix $\mathbf{\Sigma}$ in Table 4 specify the correlation structure across the three models. Specifically, it is the correlation structure across the λ_h, μ_h , and the δ_h values in Equations (4), (7), and (10), respectively. The λ_h and μ_h values can be interpreted as a measure of the base-level household-specific

**Table 3(a) Parameter Estimates
(Interpurchase-Time Model)**

Parameter	Estimate
ν_1	2.2809* (0.04801)
λ'	0.0751* (0.00482)
λ''	−0.00138* (0.000169)
$\bar{\lambda}_1$	−0.0401* (0.01218)
τ_1^2	0.0326* (0.00255)
λ_2	−0.0324 (0.03961)

**Table 3(b) Parameter Estimates
(Purchase-Amount Model)**

Parameter	Estimate
σ^2	0.2050* (0.00382)
μ'	0.0191* (0.00350)
μ''	−0.00047* (0.000121)
$\bar{\mu}_1$	−0.0015 (0.00774)
τ_2^2	0.0131* (0.00080)
μ_2	−0.0994* (0.02648)

**Table 3(c) Parameter Estimates
(Lifetime-Hazard Model)**

Parameter	Estimate
δ'	1.508* (0.08855)
δ''	−0.0682* (0.00477)
δ'''	0.00103* (0.000086)
$\bar{\delta}_1$	−0.1096 (0.05762)
τ_3^2	0.1716* (0.03388)
$\bar{\delta}_2$	0.6477* (0.09251)
τ_4^2	0.1385* (0.02428)
δ_3	−0.2317 (0.21612)

expected interpurchase times and the expected purchase amounts, respectively, whereas the δ_{hi} values can be interpreted as a measure of household-specific base-level risk of defection from the firm at each pur-

Table 4 Parameter Estimates (The Correlation Structure)

$\bar{\Omega} = \begin{pmatrix} \bar{\lambda} \\ \bar{\mu} \\ \bar{\delta} \end{pmatrix} = \begin{pmatrix} 2.2587^* \\ (0.03843) \\ 2.7021^* \\ (0.02493) \\ -9.3303^* \\ (0.41193) \end{pmatrix}$			
Σ matrix			
	$TIME_{hi}$	$\log AMNT_{hi}$	$h(LIFE_{hi})$
$TIME_{hi}$	0.2073* (0.01588)	0.0164* (0.00673)	0.8435* (0.08982)
$\log AMNT_{hi}$	0.0164* (0.00673)	0.0757* (0.00567)	0.0422 (0.03788)
$h(LIFE_{hi})$	0.8435* (0.08982)	0.0422 (0.03788)	6.1262* (0.94064)

chase occasion. The mean of the estimated distribution of these parameters is given in Table 4 ($\bar{\lambda}$, $\bar{\mu}$, and $\bar{\delta}$, respectively). For example, the estimated value of $\bar{\lambda}$ is 2.2587, which corresponds to approximately 9.5 weeks [= exp(2.2587)]. The point estimates of λ_{hi} show that 95% of the households have a base-level expected interpurchase time between 4.4 and 18.6 weeks.¹² As mentioned earlier, λ_{hi} is part of multivariate normal correlation structure (Equation (13)), the estimated parameters of which are given in Table 4 (see $\bar{\lambda}$). Similarly, the estimates of μ_{hi} correspond to a variation of \$11.1 to \$20.9 in the base-level expected purchase amounts whereas the estimates of δ_{hi} correspond to a variation of less than 0.001% to 1.3% in the base “risk” of defection across customers.

The Σ matrix (Equation (13)) in Table 4 specifies the covariance structure across these household-specific intercepts. All the estimated terms of the covariance matrix have intuitive signs. Interpurchase times and purchase amounts have a significant positive correlation, therefore, customers who tend to delay their purchases in some way “make up” by spending “more” whenever they do purchase.¹³ The correlation across interpurchase times and the risk of leaving is also positive and significant (a correlation of 75%), implying that longer spells of interpurchase times are associated with greater risk of a customer leaving the firm. The third covariance (that between purchase amounts and risk of leaving) turns out to be insignificant in our model.

We now discuss the parameter estimates of the interpurchase-time model (Table 3(a)), followed by a

¹² The estimates of household-specific parameters have not been reported in the manuscript for sake of brevity.

¹³ When the covariance matrix is converted to a correlation matrix, this correlation is found to be close to 13% [=0.0164/((0.2073 * 0.0757)^(0.5))].

discussion of the estimates of the purchase-amount model and the risk of defection model (Tables 3(b) and 3(c), respectively).

The parameters λ' and λ'' in Table 3(a) are the second-order polynomial approximation of the non-stationarity in interpurchase times after controlling for the effect of household-specific intercept λ_h and covariates used in the model (Equation (4)). The signs of these coefficients indicate that interpurchase times tend to increase and then decrease as purchase occasions progress. It is possible that as purchase occasions progress, more and more customers “try out” the service and, in the long run, the less “loyal” and the more “erratic” purchasers have left the firm, and those remaining with the firm have consistent and perhaps higher frequencies of purchase.

The parameters $(\bar{\lambda}_1, \tau_1^2)$ specify the mean and variance, respectively, of the normal heterogeneity distribution over the household-specific response parameters (λ_{1h}) representing the effect of lag interpurchase time on the current interpurchase time. These are estimated as $(-0.0401, 0.0326)$, implying that, on average, the impact of lag interpurchase time on current interpurchase time is not significant. Now when we look at the customer-specific estimates of λ_{1h} (not reported in the manuscript), we find that there are only 3.1% customers with a “significant” estimate of λ_{1h} (1.3% have a negative estimate, whereas the remaining 1.8% have a positive estimate), also indicating that the “average” effect of lagged interpurchase time on the current interpurchase time in the population is minimal (almost absent). Finally, the parameter λ_2 ($= -0.0324$) is not significantly different from 0, indicating that both male and female customers have similar interpurchase time intervals.

Now consider the estimates of the purchase amount model in Table 3(b). The parameters μ' and μ'' approximate the nonstationarity in purchase amounts. Their signs indicate that purchase amounts initially increase and then decrease across purchase occasions. The parameter μ_2 ($= -0.0994$) is significant and negative, indicating that women tend to spend less compared to men. On average, women tend to spend 9% [$= 1 - \exp(-0.0994)$] less dollars per occasion than men.

The parameters $(\bar{\mu}_1, \tau_2^2)$ specify the mean and variance, respectively, of the normal heterogeneity distribution over the household-specific response parameter for the effect of lag purchase amount on the current purchase amount (μ_{1h}) in the purchase-amount model. Their estimates of $(-0.0015, 0.0131)$ imply that, on average, the impact of lag purchase amount on current purchase amount is not significant. When we consider the customer-specific estimates of μ_{1h} , we find that there are only 1.1% of customers with a significant μ_{1h} , also indicating that the

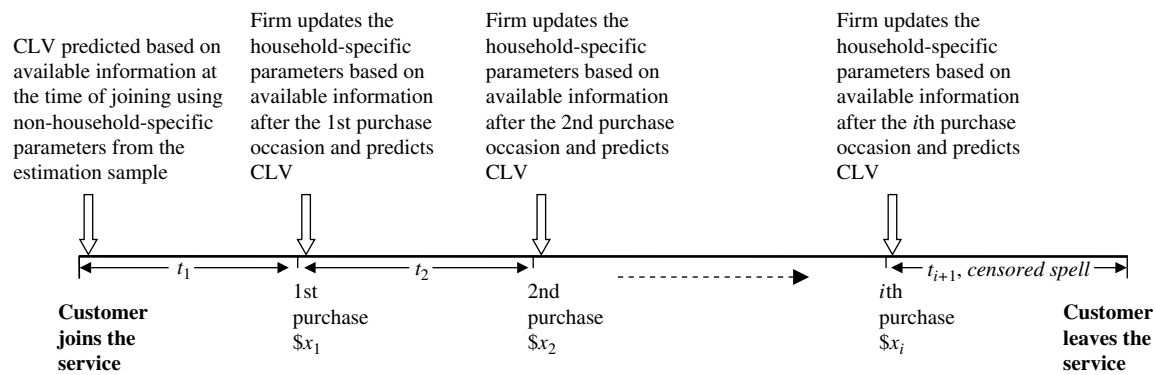
“average” effect of increases in lag purchase amounts on the current purchase amounts is minimal, i.e., insignificant.

Table 3(c) contains estimates from the customer-defection model. The parameters δ' , δ'' , and δ''' form a third-degree polynomial approximation (as mentioned earlier, the higher-order terms in the polynomial were insignificant) of the nonstationarity in the hazard rate of customers leaving the membership at each purchase occasion after controlling for the effect of household-specific intercept δ_h and covariates used in the model (Equation (10)). The signs and magnitude of the δ' , δ'' , δ''' parameters mirror the empirical hazard rate shown earlier in Figure 4(b) in that the hazard initially rises, then falls, and then rises again as purchase occasions progress. Uncovering the pattern of mortality is a very important part of the model and can be leveraged by the firm in improving predictions of customer lifetime value. To the best of our knowledge, the extant literature has not studied such a kind of application where the firm jointly uses customer mortality pattern and customer purchase behavior to better predict CLV.

The parameters $(\bar{\delta}_1, \tau_3^2)$ specify the mean and variance of the normal heterogeneity distribution over δ_{1h} 's that are the customer-specific response parameters for the effect of lag interpurchase time on the risk of defection. The estimates of $(\bar{\delta}_1, \tau_3^2)$, in other words, $(-0.1096, 0.1716)$ imply that $\bar{\delta}_1$, which is the average impact of lag interpurchase times on the risk of defection, is not significant. Alternately, looking at the customer-specific δ_{1h} values, we find that none are estimated to be significantly different from zero, implying that there is virtually no impact of lag interpurchase times on the risk of defection.

Similarly, the parameters $(\bar{\delta}_2, \tau_4^2)$ [estimated as $(0.6477, 0.1385)$] specify the mean and variance of the normal heterogeneity distribution over δ_{2h} 's that measure the impact of lag purchase amount on the risk of defection. On average, the estimates show a significant impact of lag purchase amount on the risk of defection. Looking at the customer-specific estimates, we find that most of the δ_{2h} values are positive and significant, also implying that higher spending by a customer corresponds to an increased subsequent risk of defection for the customer. The remaining parameter in Table 3(c), δ_3 , is the impact of gender on the risk of defection. The estimated value of -0.2317 is not significant, implying that men and women tend to have similar risks of defection.

In summary, the estimates show that there is significant nonstationarity in all of the three outcomes modeled (i.e., interpurchase time, amount spent, and risk of defection). Therefore, consideration of nonstationarity in measuring lifetime value of customers is likely to improve the measurements. We find that higher

Figure 6 Visual Depiction of Customer Lifetime Value Prediction for a Customer

spending by a customer is related to an increased risk of subsequent defection, and female customers spend less than male customers. The significance of correlations between the outcomes modeled shows the appropriateness of the joint modeling approach that we follow.

In the next section, we illustrate the usefulness of our model by applying it on a validation sample to predict the present values of the lifetime revenues of customers at each purchase occasion, i.e., customer lifetime value at each purchase occasion. We then compare the performance of the proposed model with some extant methods of CLV estimation and customer targeting.

5. Application of the Proposed Model

We consider two related applications of the proposed model and illustrate the usefulness of the model compared the extant methods used.¹⁴ The first application is in predicting customer lifetime values and the second application is in targeting valuable customers.

5.1. Predicting Customer Lifetime Value

We apply the proposed model to predict the present value of future customer lifetime revenues at each purchase occasion for each customer in a validation data sample. This sample consisted of 500 past customers (a total of 3,547 purchase occasions) spread across a total of 29 purchase occasions (i.e., the maximum number of times any customer bought from the firm in this validation data set was 29). Because we know the actual lifetimes of all of these 500 customers, we can test the performance of our model in predicting customer lifetime values. Figure 6 is helpful in illustrating the prediction of CLV.

At the time of membership initiation (time zero), all that the firm knows about the customer (in terms of relevance to prediction using our proposed model) is

the gender of the person. The lagged value of “time to next purchase” (interpurchase time) and the lagged value of purchase amount do not exist. So, using the gender covariate and using the non-household-specific parameters (in Tables 3(a)–3(c) and 4) the firm predicts (a) the *probability* of defecting before the first purchase “ p_1 ,” (b) the *time* to first purchase “ t_1 ,” and (c) the *amount* of the first purchase “ x_1 .” These three predicted values are then used in a simulation of the entire lifespan of the customer.

The simulation is done as follows: the probability p_1 is compared to a uniform(0, 1) draw and the “death” event before the next purchase occasion decided. If simulated “death” does not occur, the customer spends x_1 amount after time t_1 . So now, in the simulation, the customer has finished the first purchase occasion. Using the non-household-specific parameters in Tables 3(a)–3(c) and 4 and the now-available lagged values of interpurchase time and purchase amount (t_1 and x_1 , respectively) the firm predicts the triad value (p_2 , t_2 , x_2) for the next (second) purchase occasion. This simulation goes on until a simulated death event occurs, at which point the stream of simulated revenues is calculated for that customer and discounted to time zero (the time of the customer joining the service) using an annual discount rate of 12%¹⁵ (Gupta et al. 2004). This is done for all the customers in the data set and, thus, a total estimate of customer lifetime value at the time of joining service is obtained.^{16, 17}

After the *first actual* purchase event is observed by the firm for a customer, the firm has some more

¹⁵ A range of discount rates from 10% to 15% was also used; the relative performance of the model vis-à-vis other models considered does not change.

¹⁶ The simulation is done 1,000 times using the set of 500 thinned posterior draws from our MCMC chain and the CLV for each customer is averaged over these 1,000 × 500 iterations.

¹⁷ Assuming costs of servicing customers to be the same across customers and, thus, without loss of generality assuming this to be 0, the estimate of future revenues discounted to the present time can be viewed as the customer lifetime value.

¹⁴ As mentioned earlier, the context of our data is unique, and this limits the choice of extant methods for comparison.

information on the customer, namely, the time to first purchase, the amount of the first purchase, and that the customer “survived” the purchase occasion. Using this information and the non-household-specific parameters (in Tables 3(a)–3(c) and 4) as priors, the firm estimates the household-specific parameters λ_h, λ_{1h} (Equation (4)), μ_h, μ_{1h} (Equation (7)), and $\delta_h, \delta_{4h}, \delta_{5h}$ (Equation (10)). A simulation exercise is again carried out as described earlier except that now, wherever applicable, the household-specific parameters are used in the simulation, the end result being an estimate of customer lifetime value after the first purchase occasion.

A similar process is followed for each purchase occasion, updating the household-specific parameters with the available information and then simulating to predict CLV. Every interaction leads to more information about the customer and, thus, it is imperative that the firm use this information in future predictions (in our context it implies that the firm update the household-specific parameters after every interaction with the customer). The net result is that at each purchase occasion the firm gets an updated estimate of the future lifetime revenues from the customer discounted to the present time.

Note that in practice, a firm would use the model as follows. Whenever the firm carries out a predictive exercise to predict the CLV of its existing customers, it will look into its existing customer database. There would be many customers at varying points in their lifespan: some would have just joined, some would have completed their first purchase occasion, some would have completed the second purchase occasion, and so on. The firm would estimate the household-specific parameters for these customers [λ_h, λ_{1h} (Equation (4)), μ_h, μ_{1h} (Equation (7)), and $\delta_h, \delta_{4h}, \delta_{5h}$ (Equation (10))] using the available purchase history for each customer and the non-household-specific parameters (in Tables 3(a)–3(c) and 4) as priors. Using these parameters, the firm would do a simulation exercise as described earlier to estimate the CLV for each customer. The firm would repeat this exercise every time it wished to obtain an estimate of the CLV for its existing customers.

To illustrate the relative advantage of the proposed model in predicting lifetime value, we compare the lifetime value estimates from our model with the following other models: (a) the extended NBD–Pareto framework; (b) a heuristic method; and (c) two models nested within our proposed model. We explain the details of these models below.

The NBD–Pareto model (by Schmittlein et al. 1987, and later extended by Schmittlein and Peterson 1994) (Model 4 here) is a well regarded model in the literature on customer lifetime valuation (Jain and Singh 2002, Reinartz and Kumar 2000), recommended to be

applied in a noncontractual context. It has often been used as a benchmark to compare various methods of lifetime valuations (Fader et al. 2005). The underlying assumptions of the extended NBD–Pareto model (Schmittlein and Peterson 1994) are a *Poisson* purchase process for individual customers (with the *Poisson* rate distributed *gamma* across the population), an *exponential* distribution for individual customer lifetimes (with the *exponential* parameter distributed as *gamma* across the population), and a *normal* distribution for the dollar purchase amounts. Given these assumptions, Schmittlein and Peterson (1994) derive (among other things) an expression for the expected future dollar volume from a customer with a given purchase history. This can then be used to calculate the present value of future customer revenues. This is what we calculate for each customer at each purchase occasion in the model comparison.

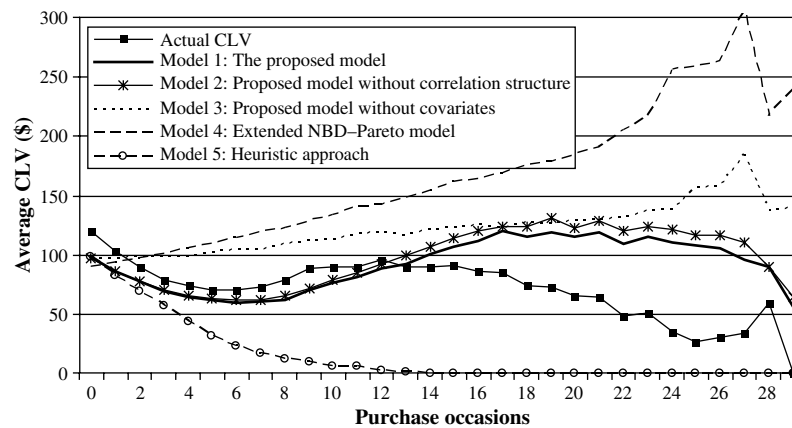
One key point in the usefulness of the NBD–Pareto framework is that the researcher *does not observe* the time when a customer becomes inactive, i.e., the end of customer lifetime with the firm. This is clearly not the case in our application where we *do* observe complete customer lifetimes. So in some sense the comparison of predictive performance of the proposed model with the extended NBD–Pareto framework may not be a direct comparison. For the sake of completeness, however, we provide a comparison with the NBD–Pareto model.

The “heuristic” model (Model 5) is a simple method whereby we take the average customer lifetime, the average interpurchase time, and the average dollar purchase amount observed in our estimation sample and use them to predict the present value of future customer revenues at each purchase occasion in our hold-out sample. The heuristic model is a simple yet useful method to calculate CLV in the absence of any available “model.”

We also compare our proposed model with two models nested in it. The first nested model is our proposed model without the correlation structure across the three components of the model, i.e., this model treats customer defection, spending, and interpurchase time as independent of each other. The second nested model is the proposed model without the covariates (including the trend parameters).

Figure 7 displays the relative predictive performance of the models. We plot the actual average customer lifetime value after each purchase occasion in our hold-out sample, and compare it with the predictive performance of the proposed model and the other models. The average customer lifetime value is the mean of the lifetime values of all customers surviving a purchase occasion. In Table 5, we report the mean absolute deviation (MAD) of the predicted lifetime value vis-à-vis the actual lifetime values for all

Figure 7 Customer Lifetime Value Predictions Across Purchase Occasions



customers across all purchase occasions. In addition, for illustration, we present the average actual and estimated customer value after the 6th purchase occasion.

The horizontal axis in Figure 7 is the purchase occasion and the vertical axis is the average lifetime value across all customers who have survived a particular purchase occasion. It is clear from Figure 7 that the proposed model outperforms the other models compared across most of the purchase occasions.

As shown in Table 5, column 3, the overall prediction from the proposed model (Model 1) is better than the other alternatives. In Figure 7, the relative advantage of the proposed model over the model without correlation (Model 2) was not visually apparent, but comparing the MAD values, we find that the proposed model does much better than the nested model without correlations across the three components (Model 2). This demonstrates that there is clear value in modeling the correlation structure because we “lose” information if we assume independence across purchase times, purchase amounts, and the risk of defection. Comparing Model 3 (the other nested model without the covariates) with Model 1, we find that Model 3 performs poorly relative to Model 1. Hence, inclusion of covariates also helps to better

predict the CLV. The MAD statistics show that the heuristic model also performs poorly relative to the proposed model.

We now compare the extended NBD-Pareto model to Model 3 (the proposed model without covariates) because the NBD-Pareto model does not include covariates. The MAD values show that Model 3 performs better than the extended NBD-Pareto model. One reason for the poor performance of extended NBD-Pareto model (relative to the proposed model) may be that it does not use the extra information in observing completed lifetimes (thus, it cannot incorporate a time-varying mortality rate), which is explicitly used in our model formulation. The trend variable δ_i (Equation (10)) included in the customer-defection model (§3.3) estimates the time-varying trend observed in customer mortality and significantly improves the prediction of customer lifetime values. This highlights the value of including the time-varying trend in the model formulation to improve CLV prediction.

5.2. Targeting Valuable Customers

In another related application of the proposed model, we apply it to “score” customers for targeting. This allows us to compare the model performance with the widely used RFM value framework. The RFM framework is a commonly used technique to score customers for a variety of purposes (e.g., targeting customers for a direct-mail campaign). As the name suggests, the RFM framework uses information on a customer’s past purchase behavior along three dimensions (recency of past purchase, frequency of past purchases, and the monetary value of past purchase) to score customers. For our analysis, we employed an “advanced form of RFM scoring” (Reinartz and Kumar 2003). We regressed the purchase amounts at each purchase occasion (in the validation data sample) on the past purchase amounts, the past interpurchase time, and the past cumulative

Table 5 Predicting Customer Lifetime Values (Comparison Across Models)

Model types	CLV (after the 6th purchase occasion) (\$)	MAD (all observations) (\$)
Actual average CLV	69.98	0
Model 1 (Proposed model)	60.04	46.93
Model 2 (Proposed model without the correlation structure)	62.22	57.64
Model 3 (Proposed model without covariates)	104.10	61.07
Model 4 (Extended NBD-Pareto model)	113.89	72.29
Model 5 (A “heuristic” approach)	23.13	61.84

Table 6 Targeting Customers (Comparison Across Models)

	Sum total CLV (\$)
Ideal baseline	295,793
Model 1 (Proposed model)	267,223
Model 2 (Proposed model without the correlation structure)	231,819
Model 3 (Proposed model without covariates)	244,916
Model 4 (Extended NBD–Pareto model)	202,873
Model 5 (A “heuristic” model)	201,168
RFM technique	197,103

frequency of purchases. Specifically, we estimated the following equations:

$$\log AMNT_{hi} \sim Normal(\psi_{hi}, \phi^2), \quad (14)$$

where ψ_{hi} is further specified as

$$\begin{aligned} \psi_{hi} = & \psi_h + \psi_1 TIME_{h,i-1} + \psi_2 FREQ_{h,t-1} \\ & + \psi_3 \log AMNT_{h,t-1}. \end{aligned} \quad (15)$$

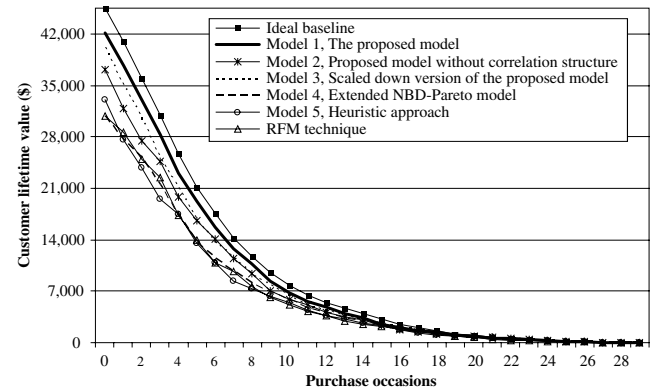
The estimated coefficients from the above equations were then used to predict the purchase amounts for the *next* purchase occasion. So, after each purchase occasion, we end up with a predicted purchase amount for the next purchase for each customer. This is used as a score for each customer after each purchase occasion. We then sorted the sample at each purchase occasion on this score and selected the top 50% of customers for targeting. The sum total of *actual* CLV of these customers was then compared at each purchase occasion with the sum total of actual CLV of similar sets of 50% of customers obtained using the proposed model and the other comparison models (Models 1–5, Table 5).^{18, 19} The results of the comparison are provided in Table 6. The table provides the sum total of CLV across all the purchase occasions for the targeted customers using the RFM technique, the proposed model, and the other models. The table also provides a similar figure for the best 50% of customers based on the actual CLV at each purchase occasion. This metric serves as an “ideal baseline” against which the performance of other techniques can be gauged.

As can be seen from Table 6, the proposed model (Model 1) outperforms the RFM technique and the other models in terms of targeting customers with the highest lifetime values. A comparison of the proposed model to the ideal baseline shows that our model is very close to the ideal baseline.

¹⁸ We also used the top 30% and 60% of customers; however, there was no significant change in the relative ranking of the various models.

¹⁹ Another method to score customers is neural nets. Such nonparametric methods might be appealing alternatives in some contexts. We thank an anonymous reviewer for pointing this out.

Figure 8 CLV of the Targeted Customers Across Purchase Occasions



Past research (Reinartz and Kumar 2003) has compared the extended NBD–Pareto model with RFM techniques and found that the extended NBD–Pareto model outperforms various RFM techniques. Our results also support this finding.

To further explore the relative advantage of various approaches in targeting customers, we plot in Figure 8 a finer version of the information contained in Table 6. We plot the sum total of CLV for the targeted customers across each of the purchase occasions in the validation data sample. The ideal baseline (the actual CLV of the top 50% of the customers) is plotted along with the CLV of the top 50% of customers using the proposed model and its variants along with the NBD–Pareto model, the heuristic approach, and the RFM technique.

Figure 8 reiterates the conclusions from Table 6 that the proposed model (and its variants) perform better in targeting customers across purchase occasions compared with using the NBD–Pareto model, the heuristic approach, or the RFM technique.

6. Summary and Discussion

Measurement of customer lifetime value is important because it is a metric in evaluating decisions in the context of customer relationship management. Because customer purchase behavior might change over time, the key drivers of CLV also might change over customer lifetime with the firm. Thus, a desirable characteristic of a measure of CLV is that it should account for past customer behavior to measure the remaining CLV at any time.

In this study, we use a hierarchical Bayes approach to model a customer’s lifetime value with the firm by explicitly accounting for her expected spending pattern over time. We estimate the model on data from a direct marketer where the purchase behavior and completed customer lifetime with the firm are observed for each customer. Furthermore, the relationship between customer lifetime and purchase behavior is not certain. Using the model estimates we

can calculate the customer lifetime value for each customer at each purchase occasion.

We compare the performance of our model in two applications on a separate validation data set. First, in measuring CLV, we compare our proposed model with the extended NBD-Pareto model, a heuristic model, and two other models nested within our proposed model. Second, in targeting customers, we compare our proposed model to all the models compared earlier and an RFM value framework. The results show that our model performs better at both predicting the customer lifetime value and targeting valuable customers than the other models. We also find that jointly modeling customer spending, interpurchase time, and the risk of customer defection, incorporating time-varying effects in the model formulation, and including relevant covariates in the model significantly improve the predictive performance of the model.

Some of our key results show that longer spells of interpurchase time are associated with a greater risk of customer leaving the firm and also larger purchase amounts (though the latter association is weak). Both male and female customers seem to have similar interpurchase time intervals; however, women spend less than men. The risk of defection is similar across male and female customers.

Most methods of estimating customer lifetime value can be best applied in specific situations where their critical assumptions are satisfied. Our approach is best suited for situations where a firm observes when a customer stops doing business with it, i.e., customer lifetimes with the firm are known to the firm after a customer leaves the firm, and customer purchase behavior is stochastic. Examples of such situations would be membership-based purchase clubs such as movie clubs, music clubs, book clubs, automobile associations, and membership-based retailers (e.g., Sams Club and Costco).

One potential drawback of this analysis may be the availability of appropriate covariates. However, the proposed model specification is flexible enough to incorporate a richer set of covariates, and thereby improve its predictive performance. What is encouraging though is that, despite limited availability of covariates, the proposed approach outperforms the extant methods of CLV prediction and customer targeting, at least in the context that is analyzed in this study.

Acknowledgments

The authors thank Joseph B. Kadane and Peter Boatwright for their valuable comments and suggestions on this paper.

All authors contributed equally. The authors' names appear in random order.

References

- Berger, P. D., N. Nasr. 1998. Customer lifetime value: Marketing models and applications. *J. Interactive Marketing* 12 17–30.
- Bhattacharya, C. B. 1998. When customers are members: Customer retention in paid membership contexts. *J. Acad. Marketing Sci.* 26(1) 31–44.
- Blattberg, R. C., J. Deighton. 1996. Manage marketing by the customer equity test. *Harvard Bus. Rev.* (July–August) 136–44.
- Boatwright, P., S. Borle, J. B. Kadane. 2003. A model of the joint distribution of purchase quantity and timing. *J. Amer. Statist. Assoc.* 98 564–572.
- Bolton, R. N. 1998. A dynamic model of the duration of the customer's relationship with a continuous service provider: The role of satisfaction. *Marketing Sci.* 17(1) 45–65.
- Borle, S., U. M. Dholakia, S. S. Singh, R. A. Westbrook. 2007. The impact of survey participation on subsequent customer behavior: An empirical investigation. *Marketing Sci.* 26(5) 711–726.
- Dowling, G. R., M. Uncles. 1997. Do customer loyalty programs really work? *Sloan Management Rev.* (Summer) 71–82.
- Fader, P. S., B. G. S. Hardie, K. L. Lee. 2005. "Counting your customers" the easy way: An alternative to the Pareto/NBD model. *Marketing Sci.* 24(2) 275–284.
- Gupta, S., D. R. Lehmann, J. A. Stuart. 2004. Valuing customers. *J. Marketing Res.* 41 7–18.
- Jain, D., S. Singh. 2002. Customer lifetime value research in marketing: A review and future directions. *J. Interactive Marketing* 16 34–46.
- Jain, D. C., N. J. Vilcassim. 1991. Investigating household purchase timing decisions: A conditional hazard function approach. *Marketing Sci.* 10(1) 1–23.
- O'Brien, L., C. Jones. 1995. Do rewards really create loyalty? *Harvard Bus. Rev.* (May–June) 75–82.
- Reinartz, W. J., V. Kumar. 2000. On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing. *J. Marketing* 64 17–35.
- Reinartz, W. J., V. Kumar. 2003. The impact of customer relationship characteristics on profitable lifetime duration. *J. Marketing* 67 77–99.
- Rossi, P. E., G. M. Allenby. 2003. Bayesian statistics and marketing. *Marketing Sci.* 22(3) 304–328.
- Rust, R., K. Lemon, V. Zeithaml. 2004. Return on marketing: Using customer equity to focus marketing strategy. *J. Marketing* 68 109–127.
- Schmittlein, D. C., R. A. Peterson. 1994. Customer base analysis: An industrial purchase process application. *Marketing Sci.* 13(1) 41–67.
- Schmittlein, D. C., D. G. Morrison, R. Colombo. 1987. Counting your customers: Who are they and what will they do next? *Management Sci.* 33(1) 1–24.
- Shmueli, G., T. P. Minka, J. B. Kadane, S. Borle, P. Boatwright. 2005. A useful distribution for fitting discrete data: Revival of the COM-Poisson. *J. Royal Statist. Soc., Ser. C* 54(1) 127–142.
- Singer, J. D., J. B. Willett. 2003. *Applied Longitudinal Data Analysis*. Oxford University Press, New York.
- Singh, S. S., D. C. Jain. 2007. Customer lifetime purchase behavior: An econometric model and empirical analysis. Working paper, Rice University, Houston, TX.
- Thomas, J. S. 2001. A methodology for linking customer acquisition to customer retention. *J. Marketing Res.* 38(May) 262–268.