CIRJE-F-962

Deriving Customer Lifetime Value from RFM Measures: Insights into Customer Retention and Acquisition

Makoto Abe
The University of Tokyo

March 2015

Deriving Customer Lifetime Value from RFM Measures:

Insights into Customer Retention and Acquisition

Makoto Abe

Graduate School of Economics
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku
Tokyo 113-0033 JAPAN
TEL & FAX: +81-3-5841-5646
E-mail: abe@e.u-tokyo.ac.jp

January 24, 2015

Deriving Customer Lifetime Value from RFM Measures:

Insights into Customer Retention and Acquisition

ABSTRACT

The wide use of RFM analysis in CRM suggests that these measures contain rather rich information about customer purchase behavior. This research, using the RFM measures of a customer, develops an individual-level CLV model that identifies the underlying behavior traits of purchase rate, lifetime and spending, which are then linked to CLV. In the application to two datasets, frequent shoppers program data from a department store and a CD chain, the model produces customer-specific metrics that are useful for identifying preferred customers and taking marketing actions targeted at the individual level in CRM. The paper then presents a retention program for existing customers that is most effective in terms of Marketing ROI, such as what action needs to be taken to which customers at which timing. For prospective customers without RFM measures, by relating the demographic characteristics to behavioral traits, insight into acquisition strategy is obtained.

Keywords: Marketing ROI, scoring models, Pareto/NBD, hierarchical Bayes, MCMC

1. INTRODUCTION

Three customer measures—recency (time since most recent purchase), frequency (number of prior purchases), and monetary value (average purchase amount per transaction), or RFM—have been used extensively in CRM with great success. At the heart of CRM is the concept of customer lifetime value (CLV), a long-term approach to identifying profitable customers and cultivating relations. The three customer measures of RFM serve the purpose well by capturing past buying behaviors parsimoniously without burdening a firm's data storage limits. Although not all firms maintain each customer's purchase history, most still accumulate customer RFM data (Hughes 2000). This suggests that rich, long-term customer information is condensed into these three measures (Buckinx and Van den Poel 2005).

Despite the importance of CLV and the usefulness of RFM measures, however, the literature provides conflicting findings on their relationship. In their predictive analysis, Malthouse and Blattberg (2005) generally found that frequency and monetary measures were positively correlated with CLV while recency was negatively correlated with CLV. Blattberg, Malthouse, and Neslin (2009), in their empirical generalization article on CLV, further state that practitioners share a similar sentiment. However, they also point out at least five academic studies that have conflicting results on the association of RFM with CLV, leading them to conclude that there is a need for further research on the issue.

Here, one must be careful to note that RFM themselves are not behavioral constructs but mere statistical measures that are observed as a result of the underlying customer behavior. In particular, recency is strongly affected by the analyst's selected time frame. For the same recency, frequent customers are more likely to be inactive than infrequent ones, thereby resulting in different CLVs. Frequency is different from purchase rate. Purchase rate considers a customer's purchases only when that customer is active. In contrast, frequency counts purchases during the entire observation period that can include a customer's inactive duration.

Hence, observed RFM measures are interrelated in a complex manner, as demonstrated by the iso-value contour plot (Fader, Hardie, and Lee 2005). A more rigorous approach to investigating the relationship between RFM and CLV is to infer the underlying behavior processes from RFM measures and then seek their association with CLV. In a non-contractual setting, appropriate behavior traits that can be derived from RFM measures as sufficient statistics are “purchase rate”, “lifetime”, and “spending per transaction”, which will be denoted as PLS throughout the paper (Fader et al. 2005).

Such model-based studies of customer behavior with PLS, however, still generate contradictory findings. Table 1 shows a partial list of such research. For instance, Reinartz and Kumar (2003) find that monetary value (monthly spending) is positively related to lifetime duration. Borle, Singh, and Jain (2008), in a discrete-time contractual setting, found a positive relationship between purchase rate and lifetime, a negative relation between purchase rate and spending per purchase, and no correlation between lifetime and spending. Singh, Borle, and Jain (2009) show that purchase rate is negatively associated with lifetime and spending per purchase, whereas lifetime and spending have a positive relationship, although the consequence of investigating only one of the two gamma parameters (the shape instead of the usual scale parameter) is unknown. Fader et al. (2005) did not find any relationship between purchase rate and average spending in their database, although they noted that its empirical generalization also requires more studies.

Table 1

So why is understanding the relationships among purchase rate, lifetime, and spending important? Borle et al. (2008), for example, found that frequent shoppers spend less per purchase. If that is the case, an obvious question is whether recruiting frequent shoppers leads to increased or decreased CLV. The answer has a potentially severe consequence for managerial decision making in CRM, where incorrect decisions could be detrimental to the firm. To assess the net influence on CLV accurately, one must tradeoff the relative magnitudes of increase in purchase rate with the decrease in spending, as well as account for any associated change in lifetime.

The objectives of this research are threefold. The first is, from the observed RFM measures of a customer, to develop an individual-level CLV model that identifies the underlying customer traits of PLS, which are then linked to CLV. By evaluating the individual difference in the intensity of PLS, one can obtain insights into their correlation and, thus, accurately assess their net impacts on CLV. The model accounts for both observed (characteristics) and unobserved customer heterogeneity. Figure 1 illustrates the framework of our approach.

Figure 1

The second objective is, using this model, to obtain normative implications for marketing programs that maximize the CLV of existing customers. In particular, we investigate implementable programs for customer retention that is most effective in terms of Marketing ROI, such as what action needs to be taken to which customers at which timing.

The third objective is, by applying our CLV model to prospective customers, to gain insight into acquisition strategy. Without transaction data, RFM measures of prospective customers do not exist and, therefore, we cannot calculate their CLV. By examining the relationship between PLS and the demographic characteristics of existing customers, we can infer the PLS of prospective customers from their demographic information alone and, hence, their CLV.

The paper is organized as follows. Section 2 discusses the general approach of our model, followed by a detailed specification that relates RFM measures to PLS, which are then linked to CLV. Then customer-specific metrics that are useful for CRM, such as expected lifetime, 1-year survival rate, etc. are derived from the model. Section 3 describes the model estimation. Section 4 presents the empirical analysis with actual customer data. Finally, Section 5 concludes the paper with summary, limitations, and future directions.

2. MODEL

2.1. The General Approach

Our approach is to construct a Pareto/NBD-based stochastic model of buyer behavior (Schmittlein, Morrison, and Colombo 1987, hereafter referred to as SMC) for estimating CLV in a "non-contractual" setting. In this case, a customer being "alive" or "dead" is inferred from recency-frequency data through simple assumptions regarding purchase behavior. The CLV research using such a Pareto/NBD model includes Fader, Hardie and Lee (2005; hereafter referred to as FHL), Reinartz and Kumar (2003), and Schimittlein and Peterson (1994; hereafter referred to as SP). Furthermore, for disaggregate modeling, we adapt the hierarchical Bayes extension of Pareto/NBD proposed by Abe (2009) to manage "heterogeneity."

Table 2

Table 2 highlights our methodology in comparison with that of SP and FHL. For recency-frequency data, both SP and FHL adopt a Pareto/NBD model that presumes a Poisson purchase and an exponential lifetime processes whose parameters are independently distributed as gamma. For monetary data, SP posit a normal-normal model, whereby purchase amounts on different occasions within a customer are normally distributed with the mean following a normal distribution in order to capture customer heterogeneity. FHL use a gamma-gamma model, whereby the normal distributions within and across customers in SP are replaced by gamma distributions. Both methodologies can be characterized as an individual-level behavior model whose parameters are compounded with a mixture distribution to capture customer heterogeneity. This in turn is estimated by an empirical Bayes method.

The proposed methodology posits the same *behavioral* assumptions as SP and FHL, but captures customer heterogeneity through a more general mixture distribution to account for the interdependence among the three behavior processes of PLS using the hierarchical Bayes (HB) framework.

Before describing the model in detail, let us summarize the reasons for our methodology.

(1) Assumptions on customer behavior are minimal and posited in various past studies.

(2) Aggregation can be managed easily without resorting to integration, and individual-level statistics can be obtained as a byproduct of MCMC estimation.

(3) It produces correct estimates of standard errors, thereby permitting accurate statistical inference even when the asymptotic assumption does not apply.

(4) It is easy to incorporate covariates.

2.2. Model Assumptions

This section describes the assumptions of the proposed HB model.

Individual Customer

Assumption 1: Poisson purchases. While active, each customer makes purchases according to a Poisson process with rate λ.

Assumption 2: Exponential lifetime. Each customer remains active for a lifetime, which has an exponentially distributed duration with dropout rate μ.

Assumption 3: Lognormal spending. Within each customer, the amounts of spending on purchase occasions are distributed as lognormal with location parameter η.

Assumptions 1 and 2 are identical to the behavioral assumptions of a Pareto/NBD model. Because their validity has been studied by other researchers (FHL; Reinartz and Kumar 2000, 2003; SMC; SP), the justification is not provided here. Assumption 3 is specified because (1) the domain of spending is positive and (2) inspection of the distributions of spending amounts within customers reveals a skewed shape resembling lognormal. As described previously, SP and FHL assume normal and gamma, respectively, to characterize the distribution of spending amounts *within* a customer.

Heterogeneity across Customers

Assumption 4: Individuals' purchase rates λ, dropout rates μ, and spending parameters η follow a multivariate lognormal distribution.

Assumption 4 permits correlation among purchase rate, lifetime, and spending parameters. Because Assumption 4 implies that $\log(\lambda)$, $\log(\mu)$, and $\log(\eta)$ follow a multivariate normal, estimation of the variance-covariance matrix is tractable using a standard Bayesian method. A Pareto/NBD combined with either a normal-normal (SP) or gamma-gamma (FHL) spending model posits independence among the three behavioral processes. We will assess the impact of this independence assumption through comparison between multivariate and independent lognormal heterogeneity. The impact of assuming a different heterogeneity shape (lognormal rather than gamma) is an empirical issue, which will be evaluated by comparing independent lognormals of the HB model with independent gammas of Pareto/NBD.

2.3. Mathematical Notations

For recency and frequency data, we will follow the standard notations $\{x, t_x, T\}$ used by SMC and FHL. Lifetime starts at time 0 (when the first transaction occurs and/or the membership starts) and customer transactions are monitored until time T. x is the number of repeat transactions observed in the time period $(0, T]$, with the last purchase (xth repeat) occurring at t_x. Hence, recency is defined as T-t_x. τ is an unobserved customer lifetime. For spending, s_n denotes the amount of spending on the nth purchase occasion of the customer under consideration. Using these mathematical notations, the previous assumptions can be expressed as follows:

(1) $$P[x \mid \lambda] = \begin{cases} \dfrac{(\lambda T)^x}{x!} e^{-\lambda T} & \text{if } \tau > T \\ \dfrac{(\lambda \tau)^x}{x!} e^{-\lambda \tau} & \text{if } \tau \le T \end{cases} \qquad x = 0,1,2,..$$

(2) $$f(\tau) = \mu e^{-\mu \tau} \qquad \tau \ge 0$$

(3) $$\log(s_n) \sim N(\log(\eta), \omega^2) \qquad s_n > 0$$

(4) $$\begin{bmatrix} \log(\lambda) \\ \log(\mu) \\ \log(\eta) \end{bmatrix} \sim MVN\left(\theta_0 = \begin{bmatrix} \theta_\lambda \\ \theta_\mu \\ \theta_\eta \end{bmatrix}, \Gamma_0 = \begin{bmatrix} \sigma_\lambda^2 & \sigma_{\lambda\mu} & \sigma_{\lambda\eta} \\ \sigma_{\mu\lambda} & \sigma_\mu^2 & \sigma_{\mu\eta} \\ \sigma_{\eta\lambda} & \sigma_{\eta\mu} & \sigma_\eta^2 \end{bmatrix} \right)$$

where N and MVN denote univariate and multivariate normal distributions, respectively. ω^2 is the variance of logarithmic spending amounts <u>*within a customer*</u>.

2.4. Expressions for Transactions, Sales, and CLV

Given the individual level parameters for (λ, μ), the expected number of transactions in arbitrary time duration w, $E[X(w)|\lambda, \mu]$, is shown by evaluating $E[\psi]$ as

(5) $$E[X(w) \mid \lambda, \mu] = \lambda E[\psi] = \frac{\lambda}{\mu}\left(1 - e^{-\mu w}\right) \qquad \text{where } \psi = \min(\tau, w).$$

The expected sales during this period w is simply the product of the expected number of transactions shown in Equation (5) and the expected spending $E[s_n|\eta, \omega]$ as

(6) $$E[sales(w) \mid \lambda, \mu, \eta, \omega] = E[s_n \mid \eta, \omega] E[X(w) \mid \lambda, \mu] = \eta e^{\omega^2/2} \frac{\lambda}{\mu}\left(1 - e^{-\mu w}\right).$$

For CLV, we define "value" to be synonymous with "revenue" because margin and cost information is unknown in this study. The general formula of CLV for an individual customer under a continuous time framework, as appropriate for a Pareto/NBD model, is expressed as

$$CLV = \int_0^\infty V(t)R(t)D(t)dt \,,$$

where $V(t)$ is the customer's value (expected revenue) at time t, $R(t)$ is the survival function (the probability that a customer remains active until at least t), and $D(t)$ is a discount factor reflecting the present value of money received at time t (FHL; Rosset et al. 2003). Translating to Assumptions 1-3, they are $V(t) = \lambda E[s_n]$ where $E[S_n] = \eta \exp(\varpi^2/2)$ from the definition of lognormal, and $R(t) = \exp(-\mu t)$. With continuously compounded discounting of an annual interest rate d, $D(t) = \exp(-\delta t)$, where $\delta = \log(1+d)$ with the time unit being a year. Therefore, our CLV reduces to the following simple expression.

(7) $$CLV = \int_0^\infty V(t)R(t)D(t)dt = \int_0^\infty \lambda \eta e^{\omega^2/2} e^{-\mu t} e^{-\delta t} dt = \frac{\lambda \eta e^{\omega^2/2}}{\mu + \delta}$$

Hence, if we can estimate λ, μ, η, and ω for each customer from RFM data, we can compute CLV as in Equation (7).

2.5. Incorporating Customer Characteristics

To control extraneous factors and gain insight into acquisition, we would like to relate customer characteristic variables for customer i, d_i (a $K \times 1$ vector) to customer specific parameters $\lambda_i, \mu_i,$ and η_i. A straightforward extension of Assumption 4 expressed in equation (4) results in a multivariate regression specification as follows:

(8) $$\begin{bmatrix} \log(\lambda_i) \\ \log(\mu_i) \\ \log(\eta_i) \end{bmatrix} \sim MVN\left(\theta_i = Bd_i, \Gamma_0 = \begin{bmatrix} \sigma_\lambda^2 & \sigma_{\lambda\mu} & \sigma_{\lambda\eta} \\ \sigma_{\mu\lambda} & \sigma_\mu^2 & \sigma_{\mu\eta} \\ \sigma_{\eta\lambda} & \sigma_{\eta\mu} & \sigma_\eta^2 \end{bmatrix} \right)$$

where B is a $3 \times K$ matrix of coefficients. When d_i contains a single element of 1 (i.e., no characteristic variables), the common mean, $\theta_0 = \theta_i$ for all customers i, is estimated.

2.6. Elasticities

Useful implications can be obtained from computing elasticities of CLV with respect to (a) λ, μ, and η, and (b) characteristic variables d_i. From equation (7),

$$E_{\lambda}^{CLV} = \frac{\partial CLV / CLV}{\partial \lambda / \lambda} = 1, \qquad E_{\mu}^{CLV} = -\frac{\mu}{\mu+\delta}, \qquad E_{\eta}^{CLV} = 1, \tag{9}$$

implying that a one percent increase in the purchase rate or spending parameter causes a one percent increase in CLV, whereas a one percent decrease in the dropout rate leads to less than a one percent increase in CLV (with the magnitude depending on the discount rate δ). Under a high interest rate, the impact of prolonging lifetime on CLV is not as rewarding because future customer value would be discounted heavily.

The effect of customer characteristics on CLV can be decomposed into purchase rate, lifetime, and spending processes to provide further managerial insight. Defining d_{ik} as the kth (continuous) characteristic of customer i, from Equations (8) and (9), the elasticity becomes

$$\begin{aligned} E_{d_{ik}}^{CLV} &= \frac{\partial CLV_i / CLV_i}{\partial d_{ik} / d_{ik}} \\ &= \left[\frac{\partial CLV_i}{\partial \lambda_i}\frac{\partial \lambda_i}{\partial d_{ik}} + \frac{\partial CLV_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial d_{ik}} + \frac{\partial CLV_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial d_{ik}} \right] \frac{d_{ik}}{CLV_i} \\ &= \left[b_{\lambda k} - \frac{b_{\mu k}\mu_i}{\mu_i + \delta} + b_{\eta k} \right] d_{ik} \\ &= Ep_{d_{ik}}^{CLV} + El_{d_{ik}}^{CLV} + Es_{d_{ik}}^{CLV} \end{aligned} \tag{10}$$

where $b_{lk}(l \in \{\lambda, \mu, \eta\}, k = 1,..., K)$ denotes $(l, k)^{th}$ element of matrix B. Applying the chain rule, the derivative with respect to d_{ik} through $\lambda_i, \mu_i,$ and η_i, results in the sum of three elasticities, $Ep_{d_{ik}}^{CLV}$, $El_{d_{ik}}^{CLV}$, and $Es_{d_{ik}}^{CLV}$ due, respectively, to purchase rate, lifetime, and spending.

3. ESTIMATION

In the previous section, simple expressions for the customer processes of purchase rate, lifetime, and spending (and thus CLV) are derived from the basic behavioral Assumptions 1, 2, and 3. To account for customer heterogeneity, the HB approach is adopted to bypass the complex aggregate expressions of the compounding mixture distribution.

The purchase rate and lifetime parts adopt the HB extension of the Pareto/NBD model proposed by Abe (2009). It is estimated by MCMC simulation through a data augmentation method. Because information about a customer being active (z = 1) or not at time T is unknown, and if not active, the dropout time ($y < T$) is also unknown, z and y are considered as latent variables.. Both z and y are randomly drawn from their posterior distributions.

The RFM data of a customer are denoted as x, t_x, T, and *as*, where x, t_x, and T are defined as in SMC, and *as* represents the average spending per purchase occasion. Without the knowledge of spending variation *within a customer* from one purchase to another, however, there is no means to infer the variance of logarithmic spending ω^2, specified in equation (3), from RFM data alone. As the RFM provides cross-sectional measures, it contains information only on spending variation *between customers*. Since it is easy to obtain ω^2 given the panel data, here we assume that ω^2 is common across customers and estimated from historical data.

Assumption 3 permits standard normal conjugate updating in Bayesian estimation, whereby the posterior mean is a precision-weighted average of the sample and the prior means. For this method to work, however, we need the mean of *log(spending)* (or equivalently, the logarithm of the geometric mean of spending amounts) from each customer, whereas the M part of RFM data provides only the arithmetic mean of spending, *as*.

Following Equation (3), $\log(s_n) \sim N(\log(\eta), \omega^2)$ implies $\mathrm{E}[s_n] = \exp(\log(\eta) + \omega^2 / 2)$. Replacing the expectation E[s_n] and log(η) by their respective sample means, $\frac{1}{x}\sum_{n=1}^{x} s_n$ and $\frac{1}{x}\sum_{n=1}^{x}\log(s_n)$, the following approximation can be obtained.

(11) $$\frac{1}{x}\sum_{n=1}^{x}\log(s_n) \cong \log(as) - \frac{1}{2}\omega^2$$

When we evaluated the accuracy of this approximation with the department store FSP data of 400 customers used in Section 4, the correlation between the actual and the approximation was 0.927, and the mean absolute percentage error was 6.8%.

3.2. Prior Specification

Reinstating customer index *i*, let us denote the customer specific parameters as $\varphi_i = [\log(\lambda_i), \log(\mu_i), \log(\eta_i)]'$, which is normally distributed with mean $\theta_i = Bd_i$ and variance-covariance matrix Γ_0 as in Equation (8). Our objective is to estimate parameters $\{\varphi_i, y_i, z_i, \forall i; B, \Gamma_0\}$ from the observed RFM data $\{x_i, t_{xi}, T_i, as_i; \forall i\}$, where the index for customer *i* is made explicit. In the HB framework, the prior of individual-level parameter φ_i corresponds to the population distribution $MVN(Bd_i, \Gamma_0)$. The priors for the hyperparameters B and Γ_0 are chosen to be multivariate normal and inverse Wishart, respectively:

$$vec(B) \sim MVN(b_{00}, \Sigma_{00}), \qquad \Gamma_0 \sim IW(\nu_{00}, \Gamma_{00})$$

These distributions are standard conjugate priors for multivariate regression models. Constants $\{b_{00}, \Sigma_{00}, \nu_{00}, \Gamma_{00}\}$ are chosen to provide very diffuse priors for the hyperparameters.

3.3. MCMC Procedure

We are now in a position to estimate parameters $\{\varphi_i, y_i, z_i, \forall i; B, \Gamma_0\}$ using an MCMC method. To estimate the joint density, we sequentially generate each parameter, given the remaining parameters, from its conditional density until convergence is achieved. Because these conditional densities are not standard distributions, the independent MH-algorithm is used.

4. EMPIRICAL ANALYSIS

4.1. Frequent Shoppers Program Data for a Department Store

We now apply the proposed model to real data. This dataset contains shopping records from 400 members of a frequent shoppers program (FSP) at a large department store in Japan.[1] These members had joined the FSP during the month of July 2000 and their transactions were recorded for 52 weeks. The first 26 weeks of data were used for model calibration and the second 26 weeks were used for validation. The available customer characteristic variables were (a) Age, (b) Gender, and (c) Food, the fraction of store visits on which food items were purchased (which is a proxy for store accessibility). The same data were also used by Abe (2009), whose descriptive statistics are reported in Table 3. The variance of *log(spending)* *within a customer*, ω^2, was estimated to be 0.895 from panel data, as discussed in Section 3.1.

Table 3

4.2. Model Validation

The MCMC steps were put through 15,000 iterations, with the last 5,000 used to construct the posterior distribution of parameters.

Table 4
----

[1] In the HB estimation, data on sample households are utilized only to construct a prior for individual customer-specific parameters. For this reason, the estimation result would be rather insensitive to the sample size, as long as it is sufficiently large (i.e., 400). One will not gain much, for example, by using a sample of 10,000 customers. Hence, the scalability is not an issue in our approach.

Table 4 shows the result of three nested HB models: Independent (variance-covariance matrix Γ_0 is diagonal without covariates), Correlated (general Γ_0 without covariates), and Full (general Γ_0 with all covariates). Correlated models with subsets of covariates are not reported here because the Full model had the best marginal log-likelihood. The performance of the Full HB model was evaluated with respect to the number of transactions and spending, obtained from Equations (5) and (6), respectively, in comparison to the benchmark Pareto/NBD-based model. The expected number of transactions, predicted by the Pareto/NBD, was multiplied by average spending as_i to come up with the customer i's spending.

Figure 2

Figure 2 shows the aggregate cumulative purchases over time. Both models provide good fit in calibration and good forecast in validation, which are separated by the vertical dashed line. With respect to the mean absolute percent errors (MAPE) between predicted and observed weekly cumulative purchases, the HB performed better for validation (1.3% vs. 1.9%) and equal to the benchmark Pareto/NBD-based approach for calibration (2.5% for both).

Figures 3 and 4

Fit statistics at the disaggregate level provide more stringent performance measures. Figure 3 shows the predicted number of transactions during the validation period, averaged across individuals and conditional on the number of purchases made during the calibration period. Figure 4 compares the predicted total spending during the validation period in a similar manner. Both figures visually demonstrate the superiority of the HB over the Pareto/NBD-based model.

Table 5

Table 5 compares the correlation and mean squared error (MSE) between prediction and observation with respect to the number of transactions and total spending at the individual customer level during the calibration and validation periods. The difference between Pareto/NBD and Independent models, aside from Empirical Bayes (EB) versus HB, can be attributed to the difference in the assumption of the heterogeneity distributions for λ and μ, whether they are independent gammas (EB) or independent lognormals (HB). The slight advantage of the Independent model over Pareto/NBD in predicting spending seems to justify the lognormal heterogeneity for this dataset. The effect of relaxing the independent assumption and incorporating the covariates is reflected, respectively, by the difference between Independent and Correlated and between Correlated and Full. Because all HB models perform similarly, the improvement in fit from accommodating correlation and covariates is minor.

In sum, the Full HB model seems to fit and predict well in comparison to the Pareto/NBD-based model, in terms of the number of transactions and spending at both the aggregate and disaggregate levels. However, the difference is minor. Considering that both models use a Bayesian method (HB vs. EB) but assume a different prior, the result seems to suggest that the estimated posterior distribution is driven mainly by data. The real advantage of the HB approach is in interpretation rather than prediction, as will be shown in the subsequent sections.

4.3. Insights into Existing Customers

4.3.1. Interpretation of the Model Estimation

Having established the validity of the HB model, let us now examine Table 4 to interpret the estimation result. Food, the fraction of store visits on which food items were purchased and a proxy for store accessibility, is the most important covariate with a significant positive coefficient for purchasing (log(λ)) and a significant negative coefficient for spending (log(η)). Managerially, food buyers tend to shop more often but spend a smaller amount on each purchase. This finding is consistent with the story told by a store manager in that, although food buyers spend a smaller amount on each shopping trip, they visit the store often enough to be considered as vital. Another significant covariate for log(η) is Age, signifying that older customers tend to spend more at each purchase.

Let us now turn our attention to the relationships among the purchase rate (λ), dropout rate (μ), and spending (η) parameters. To check whether the independence assumption of the Pareto/NBD is satisfied, the correlation of Γ_0 must be tested on the intercept-only model (Correlated) but not the covariate model (Full). The reason is that if covariates explain the correlation among λ, μ, and η completely, then no correlation remains in the error term as captured by Γ_0. First, Table 4 indicated that the correlation between log(λ) and log(μ) is not significantly different from zero, implying that the assumption of Pareto/NBD holds here. Second, the correlation between log(λ) and log(η) is significantly negative (-0.28), consistent with the Food variable having opposite signs on log(λ) and log(η) in the Full model. Figure 5 presents the scatter plot for the posterior means of the individual λ_i and η_i $(i = 1,..,400)$. One can visually observe the correlation.

Figure 5

Hence, the assumption of the independence between transaction and spending components in the Pareto/NBD-based model (SP and FHL) does not hold in this dataset. For researchers using the SP and FHL models, this finding emphasizes the importance of verifying the independence assumption (as was done in SP and FHL). Managerially, this negative correlation implies that a frequent shopper tends to spend a smaller amount on each purchase. Furthermore, this correlation remains even after accounting for differences in customer characteristics, specifically Food (store accessibility), Age, and Gender, as seen from the correlation for the Full model. No correlation was found between purchase rate and lifetime or between lifetime and spending per purchase.

4.3.2. Customer-specific Metrics and CLV

Table 6 presents nine customer-specific metrics for the top and bottom 10 customers in terms of CLV, along with the average, minimum, and maximum for the entire sample of 400 customers: posterior means of λ_i, $\mu_{i,}$, and η_i; expected lifetime; survival rate after one year; probability of being active at the end of the calibration period; an expected number of transactions (using equation (5)); expected total spending during the validation period (using equation (6)); and CLV (using equation (7)). In computing CLV, an annual interest rate of 15% (δ = 0.0027 per week) was assumed as in FHL.

Table 6

There exists much heterogeneity across customers despite the use of the Bayesian shrinkage estimation. The mean expected lifetime is 10.0 years with the maximum and minimum of 24.7 and 1.3 years, respectively. The probability of being active at the end of the calibration period ranges from 0.18 to 1.0, with the average being 0.93. Over the validation period of 26 weeks, the expected number of transactions was 16.0 times with the total amount of 74,000 yen on average (divide by 100 to convert to the approximate US dollars). CLV ranges from 40,000 yen to 10.2 million yen, with an approximate average of 0.69 million yen.

Figure 6

Figure 6 shows a gain chart (solid line) in which customers are sorted according to the decreasing order of CLV, and the cumulative CLV (y-axis, where the total CLV is normalized to 1) is plotted against the number of customers (x-axis). In addition, two gain charts are plotted. The dash-dotted line is based solely on recency criterion, whereby customers are sorted in the order of increasing recency (from most recent to least recent). The dotted line is a gain chart based on customers being ordered according to the sum of the three rankings of recency, frequency, and monetary value. The 45 degree dashed line corresponds to the cumulative CLV for randomly ordered customers. This figure implies that recency criterion alone is not sufficient to identify good customers, although many companies use this criterion. On the other hand, combined use of the three measures (recency, frequency, and monetary value), even with the naïve equal weighting scheme, seems to provide a rather accurate ordering of CLV. This finding strongly supports the wide use of RFM analysis and regression-type scoring models among practitioners for identifying good customers (Malthouse and Blattberg 2005). While RFM measures can produce the relative ranking of customers well, without the proposed model, the absolute CLV figure itself cannot be obtained.

4.3.3. Customer Base and Customer Equity

We can compute the expected number of active customers (customer base) at the end of the calibration period, that is January 1, 2001, from Table 6. Customer base is the sum of active probabilities (column 7) for all customers. Although the dataset contains 400 customers, the effective number of active customers on January 1, 2001 is only 371.6 ($= 400 \times 0.929$). When a customer turnover is high, customer base becomes much smaller than the registered number of customers in the dataset.

Customer equity is the expected CLV generated by all active customers at that time. By aggregating the products of a CLV and the active probability on January 1, 2001, for all 400 customers, it becomes 273 million yen. Customer base and customer equity are firm’s valuable long-term indicators, neither of which are provided by the accounting statement.

4.3.4. Retention Program for Existing Customers

Let us assume that, if a customer were indeed inactive, an investment level of *c* per customer would win her back with probability *r(c)*. Such a response function *r(c)*, as shown in Figure 7, can be constructed by decision calculus (Blattberg and Deighton 1996; Little 1971). The figure implies that, by sending a coupon valued at 1,000 yen (i.e., investment *c=1,000*), there is a 50% chance that inactive customers can be brought back. However, no matter how large the coupon's value is, the firm cannot bring back inactive customers with a probability of more than 0.8. The effect of such a coupon on the change in CLV is derived in the appendix. We now consider two examples of customer-specific coupon mailing.

Figure 7

Customer Retention Example 1: Given that a coupon with a different value is sent to each customer on January 1, 2001, what level of *c maximizes the increase of her CLV?**

The second column of Table 7 presents *c** for the same 20 customers as in Table 6 (i.e., top 10 and bottom 10 in terms of CLV). The optimum coupon value varied by customers between 0 yen and 1,500 yen, and the average was 500 yen. For 74 out of 400 customers, *c**=0. This implies that, for these customers, a coupon increases their CLV only by a small amount such that the investment cannot be recouped. The optimal formula for *c**, (A3) in the appendix, indicates that a higher coupon value is more effective for customers with a higher CLV and a lower active probability, which is consistent with the general pattern of Table 7.

Table 7

Customer Retention Example 2: To increase CLV by 10,000 yen, how many non-purchase days should a firm wait before mailing a coupon with a face value of 500 yen?

Here, the logic is as follows. If an active probability is sufficiently high, the investment would be wasted (as in (A3)). An active probability decreases as time passes (as in (A1)). But if the firm waited too long, an investment of 500 yen would not be sufficient to induce an increase of CLV by 10,000 yen. Therefore, there is an optimum timing to step in with the coupon. The third column of Table 7 shows this timing (i.e., non-purchase days or recency), which varies by customers from 3.5 days to 120.4 days. In general, an earlier action is better for customers with a higher CLV. The pattern is consistent with the formula (A4) in the appendix.

4.4. Insights into Prospective Customers

We now focus our attention to prospective customers for the purpose of acquisition. In particular, to assess the influence of demographic covariates on CLV, we examine the elasticity of these covariates on PLS and CLV.

4.4.1. Elasticity of Covariates on CLV

Table 8 reports the decomposition of the elasticity of CLV with respect to each covariate into purchase rate, lifetime, and spending components, as shown in Equation (10). To account for parameter uncertainty, elasticity is computed for each set of the 5000 MCMC draws of b_{lk} and μ_i according to Equation (10). This is then averaged over the 5000 draws and 400 customers. When the posterior mean of b_{lk} and μ_i is directly substituted into Equation (10) (bottom table) instead of averaging over MCMC draws (top table), elasticity with respect to the lifetime component is overestimated by about 50% (because of nonlinearity in μ_i). This overestimation occurs even when customer heterogeneity is accounted for.

One of the advantages of our Bayesian approach is that parameter uncertainty can be evaluated easily with the sampling estimation method. Ignoring their uncertainty and computing various statistics from their point-estimates, say MLE, as if parameters are deterministic, could produce a biased result and lead to incorrect managerial decisions (Gupta 1988).

Table 8

4.4.2. Implication for New Customer Acquisition

To clarify the impact of covariates, the solid line in Figure 8 plots the value of log(CLV) for different values of a covariate when the remaining two covariates are fixed at their mean values. These graphs are computed using the mean estimate of the coefficients of the Full model shown in Table 4, assuming that all covariates are continuous. For the Female covariate, therefore, it should be interpreted as how log(CLV) varies when the gender mixture is changed from the current level of 93.3% female, while keeping the other two covariates unchanged. The dotted vertical line indicates the mean value of the covariate under consideration. Both Food and Age have strong influences on log(CLV), whereas Female exerts a very weak influence, consistent with the significance of these covariates.

Figure 8

Figure 8 also attempts to decompose the influence of covariates on log(CLV) into three components: purchase rate, lifetime, and spending. Taking the logarithm of the basic formula of CLV in Equation (7) results in the following summation expression:

$$\log(CLV) = -\log(\mu+\delta) + \log(\lambda) + \log(\eta) + \omega^2/2$$

$$= [\text{Lifetime } \mu \text{ component}] + [\text{Purchase rate } \lambda \text{ component}] + [\text{Spending } \eta \text{ component}] + \text{constant}$$

The graph can be interpreted as stacking the lifetime, purchase rate, and spending components from top to bottom, thus constituting the overall log(CLV). To account for the scale differences among these components, each was adjusted to become 1 at the mean value of the covariate. Therefore, log(CLV) = 3 at the dotted vertical line.

The direction and magnitude of the effect of each covariate on the three components are consistent with the signs of the posterior means $b_{lk}(l \in \{\lambda, \mu, \eta\}, k = 1,..., K)$. Increasing the fraction of food buyers improves lifetime and purchase rate, but decreases spending per purchase, yielding a net increase in the overall CLV. Increasing the fraction of elderly people increases the spending without much influence on lifetime and purchase rate, thereby resulting in a net increase in the overall CLV. Increasing the fraction of females leads to little improvement in all three components and, hence, a negligible increase in the overall CLV.

Elasticity decomposition provides managers with useful insights into acquisition. An effort to manipulate certain customer characteristics might impact lifetime, purchase rate, and spending components in opposite directions, thereby canceling each other to produce a collectively reduced effect as the total on CLV. For example, much of the improvement in purchase rate (from increasing the fraction of food buyers) is negated by the decline in spending per purchase. In addition, only the lifetime improvement provides the net contribution to CLV, as can be seen from Table 8 and the near flat dashed line of Figure 8. In contrast, an effort to increase the proportion of elderly people is met with a boost in CLV, due to increased spending per purchase with only a small negative influence on purchase rate.

To build an effective acquisition strategy from these results, managers must strike a fine balance between desired customer characteristics (i.e., demographics), desired behavioral profiles (i.e., purchase rate, lifetime, and spending), responsiveness (elasticity) of the characteristic covariates on CLV, and the acquisition cost of the desired target customers.

4.5. Second Dataset: Retail FSP Data for a Music CD Chain

The second dataset, which was also analyzed in FHL, contains the record of 500 customers in an FSP of a large music CD store chain. The period covers 52 weeks beginning September 2003. The following customer characteristics were used as covariates: the amount of the initial purchase, age, and gender.

Table 9

Table 9 reports the model estimation. Let us first examine the Full HB model, which results in the highest marginal log-likelihood, for significant explanatory variables. First, the amount of an initial purchase is positively significant on log(λ) and log(η), implying that customers with a larger initial purchase tend to buy more often and spend more per purchase in subsequent purchases. Second, older customers appear to spend more per purchase.

Next, we turn our attention to the Correlated model for the relationship among λ, μ, and η. First, we see that the correlation between log(λ) and log(μ) is not significantly different from 0, implying that the assumption of Pareto/NBD holds here. Second, the correlation between log(λ) and log(η) is significantly positive (0.14), consistent with the initial purchase covariate having the same signs on log(λ) and log(η). Once again, the independence assumption of the transaction and spending components does not hold here. This time, however, the sign is in the opposite direction, implying that the correlation between purchase rate and spending per occasion is data dependent. Managerially, the correlation implies that frequent buyers spend more per shopping occasion. Also, note that when covariates are included (Full model), the correlation is no longer significant. This fact indicates that the difference in initial amount and age can explain the correlation between purchase rate and spending.

Table 10 and Figure 9

Table 10 shows the elasticity decomposition of CLV into purchase rate, lifetime, and spending components. When parameter uncertainty is not accounted for, the lifetime component is overestimated by approximately 20%, as was the case for the department store data. The elasticity decomposition of log(CLV) into the three components for varying levels of the three covariates is presented in Figure 9. A higher initial purchase amount is related to a higher CLV by increasing the purchase rate and spending with almost no change in lifetime. Older customers are associated with a lower purchase rate, longer lifetime, and higher spending per purchase with a positive net contribution to CLV. Female customers are associated with a lower purchase rate, shorter lifetime, and less spending with a negative net contribution to CLV.

5. CONCLUSIONS

The wide use of RFM analysis in CRM suggests that these measures contain rather rich information about customer purchase behavior. However, the existing literature provides conflicting findings on the relation between RFM and CLV, and several authors have advocated the need for further studies to provide empirical generalization (Blattberg, Malthouse, and Neslin 2009). The present research sought to clarify the issue through the identification of the underlying customer traits characterized by the interrelated behavior process of purchase rate, lifetime, and spending from statistical RFM measures. The PLS process posited the same behavioral assumptions as the established Pareto/NBD-based model studied by other researchers. The hierarchical Bayesian extension for constructing the individual-level CLV model permitted the application of accurate statistical inference on correlations, while controlling covariates and avoiding complex aggregation. The model also related customer characteristics to the buyer behaviors of purchase rate, lifetime, and spending, which were, in turn, linked to CLV to provide useful insights into retention of existing customers as well as acquisition of new customers.

Two FSP datasets, one from a department store and another from a CD chain, were investigated in the empirical analysis. The proposed CLV model provided nine customer-specific metrics: posterior means of λ_i, $\mu_{i,}$, and η_i; expected lifetime; survival rate after one year; probability of being active at the end of the calibration period; an expected number of transactions; expected total spending during the validation period; and CLV. These metrics are especially useful for identifying preferred customers and taking marketing actions targeted at the individual level in CRM. By maximizing Marketing ROI, we illustrated two examples of retention couponing that specifies what value needs to be mailed to which customers at which timing.

It was also found that recency criterion alone was not sufficient to identify good customers, although many companies use this criterion. On the other hand, combined use of the three measures (recency, frequency, and monetary value), even with the naïve equal weighting scheme, seemed to provide a rather accurate ordering of CLV. This finding strongly supports the wide use of RFM analysis and regression-type scoring models among practitioners in identifying good customers.

Finally, by relating the behavioral traits of PLS with demographic characteristics, we obtained insights into acquisition strategy for prospective customers with a high CLV. For example, the first dataset exhibited a statistically significant negative correlation between purchase rate and spending (-0.28). In such a case, recruiting food buyers for the purpose of improving purchase rate would be negated by the decline in spending per purchase and, as a result, only an improvement in lifetime contributed to the net increase in CLV. Note that a correlation between purchase rate and spending is data dependent. Specifically, one of our datasets exhibited a negative relation, whereas the other dataset showed a positive relation. For a negative relationship, their relative magnitudes must be evaluated to assess the net impact on CLV --- an insight that is crucial, especially in the context of customer acquisition.

The current study is only the beginning of a stream of research addressing customer behavior in a “non-contractual” setting. Possible extensions are synonymous with limitations of the proposed method. From the consideration of RFM measures as sufficient statistics, the current model posited the behavioral assumptions of a Poisson purchase, an exponential lifetime and lognormal spending. With a customer's complete transaction history, however, more elaborate behavioral phenomena can be modeled.

ACKNOWLEDGMENTS

I would like to express my sincere thanks to Prof. John D. C. Little, who supported me throughout my career since I was a student at MIT. I am also deeply grateful to Glen Urban and John Hauser, who served my doctoral dissertation committee, for editing this book. This research was supported by grant #19530383 from the Japan Society for the Promotion of Science.

APPENDIX: Derivation for the Effect of a Retention Program on CLV

Following Blattberg and Deighton (1996), if a customer were indeed inactive, an investment level of *c* per customer would win her back with probability *r(c)*, where

$$r(c) = R\left(1 - e^{-kc}\right) \qquad \text{where } c \geq 0 \ .$$

Here *R* and *k* are parameters, which can be estimated by decision calculus. Because whether a customer is active is unknown, we use its stochastic metric, the predicted probability of being active *p* (Table 6), derived from the proposed model, as shown below (Abe 2009).

$$\text{(A1)} \qquad p = \frac{1}{1 + \frac{\mu}{\lambda+\mu}\left[e^{(\lambda+\mu)(recency)} - 1\right]}$$

Let us denote the probability of being active with and without the retention program, respectively, as p_a and *p*. Then, by assumption, $p_a = p + (1-p)r(c)$. Therefore, an increase in CLV as a result of the program, *Δ*, is expressed as Equation (A2).

$$\text{(A2)} \qquad \begin{aligned} \Delta &= \left(p_a V - c\right) - pV \\ &= (1-p)r(c)V - c \qquad \text{where } V = CLV \end{aligned}$$

Customer Retention Example 1: Given that a coupon with a different value is sent to each customer on January 1, 2001, what level of *c maximizes the increases of her CLV?**

The optimum *c** is obtained by maximizing *Δ* with respect to *c* by solving the first-order condition of (A2) as

$$\text{(A3)} \qquad c^* = \frac{1}{k}\ln(1-p)kRV \ .$$

Customer Retention Example 2: To increase CLV by 10,000 yen, how many non-purchase days should a firm wait before mailing a coupon with a face value of 500 yen?

Substitute *Δ*=10000 and *c*=500, and solve (A2) with respect to active probability *p**.

$$p^* = 1 - \frac{c + \Delta}{r(c)V}$$

Substitute *p*=*p** in (A1) and solve for *recency*.

$$\text{(A4)} \qquad recency^* = \frac{1}{\lambda+\mu}\ln\left[\frac{\lambda+\mu}{\mu}\frac{c+\Delta}{r(c)V - c - \Delta} + 1\right]$$

REFERENCES

Abe, Makoto (2009), "'Counting Your Customers' One By One: A Hierarchical Bayes Extension to the Pareto/NBD Model," *Marketing Science*, 28 (3), 541-553.

Blattberg, Robert C. and John Deighton (1996), "Manage Marketing by the Customer Equity Test," *Harvard Business Review*, (July-August), 36-44.

Blattberg, Robert C., Edward C. Malthouse and Scott Neslin (2009), "Lifetime Value: Empirical Generalizations and Some Conceptual Questions," *Journal of Interactive Marketing*, 23 (2), 157-168.

Borle, Sharad, Siddharth S. Singh and Dipak C. Jain (2008), "Customer Lifetime Value Measurement," *Management Science*, 54 (1), 100-112.

Buckinx, Wouter and Dirk Van den Poel (2005), "Customer Base Analysis: Partial Defection of Behaviourally Loyal Clients in a Non-Contractual FMCG Retail Setting," *European Journal of Operational Research*, 164 (1), 252-268.

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2005), "RFM and CLV: Using Iso-Value Curves for Customer Base Analysis," *Journal of Marketing Research*, 42 (4), 415-430.

Gupta, Sunil (1988), "Impact of Sales Promotions on When, What, and How Much to Buy," *Journal of Marketing Research*, 25 (4), 342-355.

Hughes, Arthur (2000), *Strategic Database Marketing* (2nd Ed.), New York: McGraw-Hill.

Little, John D.C. (1971), "Models and Managers: The Concept of a Decision Calculus," *Management Science*, 16 (8) (1971), pp. B466–B485

Malthouse, Edward C. and Robert C. Blattberg (2005), "Can We Predict Customer Lifetime Value?" *Journal of Interactive Marketing*, 19 (1), 2-15.

Reinartz, Werner J. 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 (4), 17-35.

----- and ----- (2003), “The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration,” *Journal of Marketing*, 67 (1), 77-99.

-----, J. S. Thomas and V. Kumar (2005), “Balancing Acquisition and Retention Resources to Maximize Customer Profitability,” *Journal of Marketing*, 69 (1), 63-79.

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., Donald G. Morrison, and Richard Colombo (1987), “Counting your customers: Who are they and what will they do next?” *Management Science*, 33 (1), 1-24.

----- and Robert A. Peterson (1994), “Customer Base Analysis: An Industrial Purchase Process Application,” *Marketing Science*, 13 (1), 41-67.

Singh, Siddharth S., Sharad Borle and Dipak C. Jain (2009), “A Generalized Framework for Estimating Customer Lifetime Value When Customer Lifetimes are not Observed,” *Quantitative Marketing and Economics*, 7 (2), 181-205.

Table 1. Conflicting Findings on the Correlation among Purchase Rate, Lifetime, and Spending

	Purchase Rate and Lifetime	Purchase Rate and Spending	Lifetime and Spending
Schmittlein & Peterson (1994)	N/A	0	+
Reinartz & Kumar (2003)	N/A	N/A	+
Reinartz et al. (2005)	+	N/A	N/A
Fader et al. (2005)	0	0	N/A
Borle et al. (2008)	+	-	0
Singh et al. (2009)	-	-	+

Table 2. Comparison with Existing Methods

Empirical Bayes Model

Data	**Model**	**Individual Behavior**	**Heterogeneity Distribution**
RF (recency-frequency)	Pareto/NBD (SMC 1987)	Poisson purchase (λ) Memoryless dropout (μ)	λ ~ Gamma μ ~ Gamma λ and μ independent
M (monetary)	normal-normal (SP 1994)	Normal spending (mean θ)	θ ~ Normal θ, λ, μ independent
	gamma-gamma (FHL 2005)	Gamma spending (scale ν)	ν ~ Gamma ν, λ, μ independent

Proposed Hierarchical Bayes Model

Data	**Model**	**Individual Behavior**	**Heterogeneity Distribution**
RF (recency-frequency)	Poisson/exponential (Abe 2009)	Poisson purchase (λ) Memoryless dropout (μ)	λ, μ, η ~ MVL λ, μ, η correlated
M (monetary)	lognormal-lognormal	Lognormal spending (location η)	

Table 3. Descriptive Statistics for the Department Store Data

	mean	std. dev.	minimum	maximum
Number of repeat purchases	16.02	16.79	0	101
Length of observation T (days)	171.24	8.81	151	181
Recency (T-t) (days)	24.94	42.82	0	181
Average purchase amount (×10^5 yen)	0.067	0.120	0.0022	1.830
Food	0.79	0.273	0	1
Age	52.7	14.6	22	87
Female	0.93	0.25	0	1

Table 4. Estimation Results of HB Models (department store)

(Figures in parentheses indicate the 2.5 and 97.5 percentiles)
*** indicates significance at the 5% level**

		Independent	Correlated	Full
Purchase rate log(λ)	**Intercept**	-0.82 (-0.93, -0.72)	-0.81 (-0.92, -0.71)	-2.03 (-2.52, -1.51)
	Food	---	---	1.50* (1.11, 1.89)
	Age	---	---	-0.21 (-0.84, 0.40)
	Female (male = 0)	---	---	0.15 (-0.20, 0.48)
Dropout Rate log(μ)	**Intercept**	-6.24 (-7.03, -5.52)	-6.13 (-7.10, -5.56)	-5.03 (-6.49, -3.57)
	Food	---	---	-1.09 (-2.66, 0.26)
	Age	---	---	-0.34 (-2.35, 1.49)
	Female (male = 0)	---	---	0.01 (-1.20, 1.38)
Spending Parameter log(η)	**Intercept**	-3.59 (-3.67, -3.51)	-3.57 (-3.64, -3.49)	-3.23 (-3.61, -2.86)
	Food	---	---	-1.34* (-1.62, -1.06)
	Age	---	---	1.18* (0.71, 1.65)
	Female (male = 0)	---	---	0.11 (-0.15, 0.39)
correlation(log(λ), log(μ))		---	-0.33 (-0.59, 0.01)	-0.24 (-0.51, 0.09)
correlation(log(λ), log(η))		---	-0.28* (-0.39, -0.17)	-0.14* (-0.25, -0.01)
correlation(log(μ), log(η))		---	-0.01 (-0.31, 0.27)	-0.07 (-0.35, 0.24)
marginal log-likelihood		-2111	-2105	-2078

Table 5. Disaggregate Fit of Pareto/NBD and HB Models (department store)

		Pareto/NBD	Independent	Correlated	Full
Spending					
Correlation	**validation**	0.80	0.83	0.83	0.83
	calibration	0.99	0.99	0.99	0.99
MSE	**validation**	0.39	0.35	0.35	0.35
	calibration	0.02	0.06	0.06	0.06
Transactions					
Correlation	**validation**	0.90	0.90	0.90	0.90
	calibration	1.00	1.00	1.00	1.00
MSE	**validation**	57.7	57.1	57.0	56.5
	calibration	1.22	4.61	4.06	3.92

Table 6. Customer-Specific Metrics for Top and Bottom 10 Customers (department store)

ID	mean(λ)	mean(μ)	mean(η)	Mean Expected lifetime (years)	1 year Survival rate	Probability of being active at the end of calibration	Expected number of transactions in validation period	Expected total spending in val. period ($\times 10^5$ yen)	CLV ($\times 10^5$ yen)
1	3.20	0.00165	0.075	24.7	0.926	1.000	81.5	9.61	102.0
2	1.98	0.00173	0.056	21.5	0.922	1.000	50.3	4.38	45.4
3	1.15	0.00205	0.097	19.9	0.910	0.999	29.2	4.43	45.1
4	2.55	0.00188	0.036	22.1	0.918	1.000	64.7	3.60	37.2
5	1.11	0.00338	0.088	11.3	0.862	0.997	27.7	3.80	33.9
6	2.23	0.00191	0.034	22.4	0.916	1.000	56.7	3.01	31.0
7	2.86	0.00206	0.027	18.6	0.910	0.999	72.5	3.01	30.3
8	1.10	0.00202	0.067	19.3	0.910	0.996	27.8	2.93	29.6
9	2.19	0.00206	0.034	19.7	0.909	1.000	55.5	2.91	29.3
10	0.87	0.00273	0.090	13.8	0.886	0.999	21.9	3.09	29.0
…	…	…	…	…	…	…	…	…	…
391	0.29	0.01218	0.011	3.3	0.665	0.379	1.7	0.03	0.6
392	0.15	0.00750	0.016	5.7	0.754	0.803	2.7	0.07	0.6
393	0.29	0.01151	0.011	3.4	0.666	0.381	1.8	0.03	0.6
394	0.10	0.03915	0.049	1.4	0.463	0.436	0.7	0.05	0.6
395	0.38	0.02974	0.009	2.1	0.555	0.182	0.5	0.01	0.6
396	0.10	0.04307	0.044	1.4	0.480	0.450	0.7	0.05	0.5
397	0.24	0.00586	0.008	6.5	0.786	0.951	5.5	0.07	0.5
398	0.20	0.00699	0.009	6.1	0.762	0.862	4.1	0.06	0.5
399	0.10	0.04713	0.034	1.3	0.454	0.420	0.7	0.04	0.4
400	0.14	0.01709	0.016	2.2	0.581	0.601	1.6	0.04	0.4
ave	0.66	0.00564	0.038	10.0	0.823	0.929	16.0	0.74	6.9
min	0.07	0.00165	0.007	1.3	0.454	0.182	0.5	0.01	0.4
max	3.78	0.04713	0.207	24.7	0.926	1.000	96.1	9.61	102.0

Table 7. Customer-Specific Retention Action for Top and Bottom 10 Customers (department store)

ID	Coupon Value (yen)	Recency to Wait (days)	Probability of being active at the end of calibration	CLV ($x10^5$ yen)
1	407	3.5	1.000	102.0
2	235	6.5	1.000	45.4
3	510	7.8	0.999	45.1
4	221	5.9	1.000	37.2
5	623	7.0	0.997	33.9
6	0	6.9	1.000	31.0
7	385	5.8	0.999	30.3
8	661	10.1	0.996	29.6
9	186	6.9	1.000	29.3
10	372	9.6	0.999	29.0
391	916	60.5	0.379	0.6
392	684	110.8	0.803	0.6
393	911	62.6	0.381	0.6
394	887	54.0	0.436	0.6
395	949	37.8	0.182	0.6
396	864	55.2	0.450	0.5
397	372	96.3	0.951	0.5
398	561	106.7	0.862	0.5
399	822	66.2	0.420	0.4
400	713	120.4	0.601	0.4
ave	500	32.6	0.929	6.9
min	0	3.5	0.182	0.4
max	1500	120.4	1.000	102.0

Table 8. Decomposition of CLV Elasticity into Three Components (department store)

Accounting for Parameter Uncertainty

	FOOD	**AGE**	**FEMALE**
Total	0.53	0.63	0.21
purchase rate: Ep^{CLV}	1.17	-0.11	0.14
lifetime: El^{CLV}	0.41	0.12	-0.04
spending: Es^{CLV}	-1.05	0.63	0.10

Ignoring Parameter Uncertainty

	FOOD	**AGE**	**FEMALE**
Total	0.76	0.69	0.18
purchase rate: Ep^{CLV}	1.17	-0.11	0.14
lifetime: El^{CLV}	0.63	0.17	-0.06
spending: Es^{CLV}	-1.05	0.63	0.10

* Note that elasticity for lifetime but neither purchase frequency nor spending is different when uncertainty is ignored. This is because, as shown in equation (10), only μ enters the elasticity formula in a nonlinear fashion.

Table 9. Estimation Results of Various Models (music CD chain)

		Independent	Correlated	Full
Purchase rate **log(λ)**	**Intercept**	-2.11 (-2.19, -2.03)	-2.11 (-2.19, -2.03)	-2.10 (-2.34, -1.85)
	Initial amount	---	---	0.37* (0.11, 0.63)
	Age	---	---	-0.26 (-0.87, 0.34)
	Female (male=0)	---	---	-0.13 (-0.29, 0.03)
Dropout Rate **log(μ)**	**Intercept**	-5.18 (-5.63, -4.74)	-5.14 (-5.64, -4.72)	-5.06 (-5.89, -4.34)
	Initial amount	---	---	0.02 (-1.09, 0.94)
	Age	---	---	-0.15 (-1.84, 1.39)
	Female (male=0)	---	---	0.05 (-0.60, 0.64)
Spending Parameter **log(η)**	**Intercept**	-1.18 (-1.22, -1.13)	-1.18 (-1.22, -1.13)	-1.49 (-1.63, -1.35)
	Initial amount	---	---	0.50* (0.36, 0.65)
	Age	---	---	0.47* (0.12, 0.82)
	Female (male=0)	---	---	-0.03 (-0.10, 0.05)
correlation **(log(λ), log(μ))**		---	0.20 (-0.02, 0.43)	0.19 (-0.04, 0.42)
correlation **(log(λ), log(η))**		---	0.14* (0.01, 0.27)	0.10 (-0.05, 0.24)
correlation **(log(μ), log(η))**		---	0.01 (-0.22, 0.24)	0.01 (-0.20, 0.22)
marginal log-likelihood		-2908	-2906	-2889

*** indicates significance at the 5% level**

Table 10. Decomposition of CLV Elasticity into Three Components (music CD chain)

	Initial Amount	**AGE**	**FEMALE**
Total	0.31	0.12	-0.10
purchase rate: Ep^{CLV}	0.13	-0.09	-0.06
lifetime: El^{CLV}	0.00	0.06	-0.02
spending: Es^{CLV}	0.18	0.15	-0.01

Figure 1. Framework of Our Approach

Underlying customer traits are derived from the observed RFM measures, which are in turn linked to CLV. The model accounts for both observed (characteristics) and unobserved customer heterogeneity.

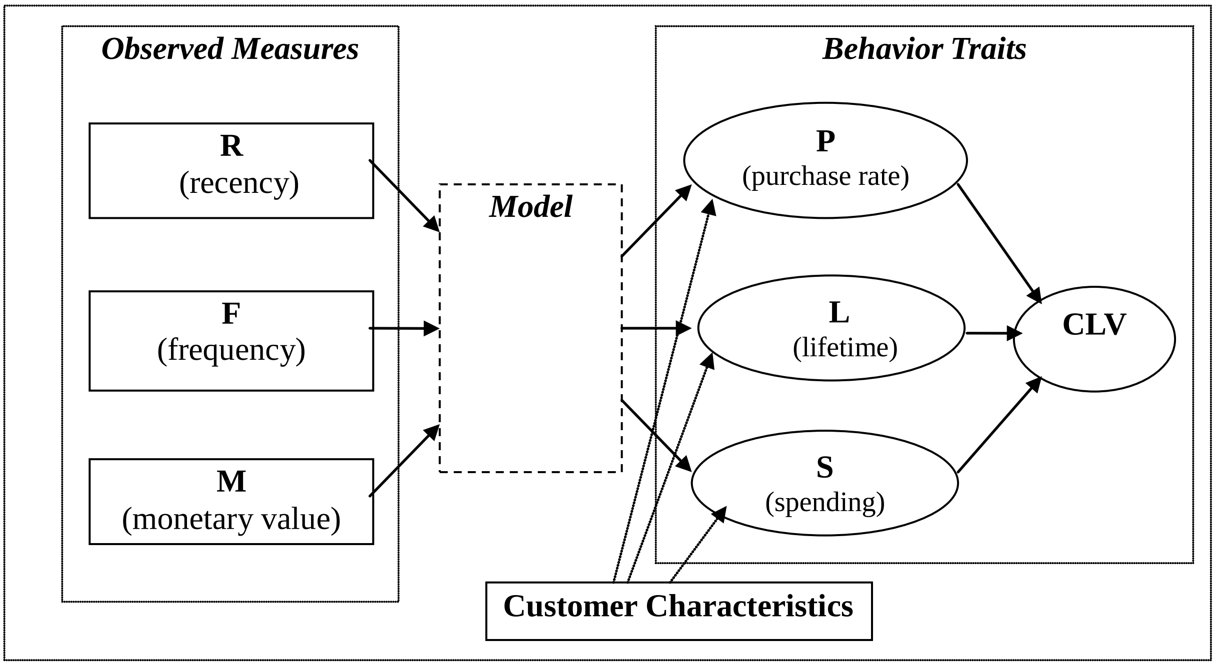


Figure 2. Weekly Cumulative Repeat Transaction Plot

The vertical dashed line separate the calibration period and forecast period. Both proposed model (HB) and Pareto/NBD models provide good fit on the aggregate cumulative purchases.

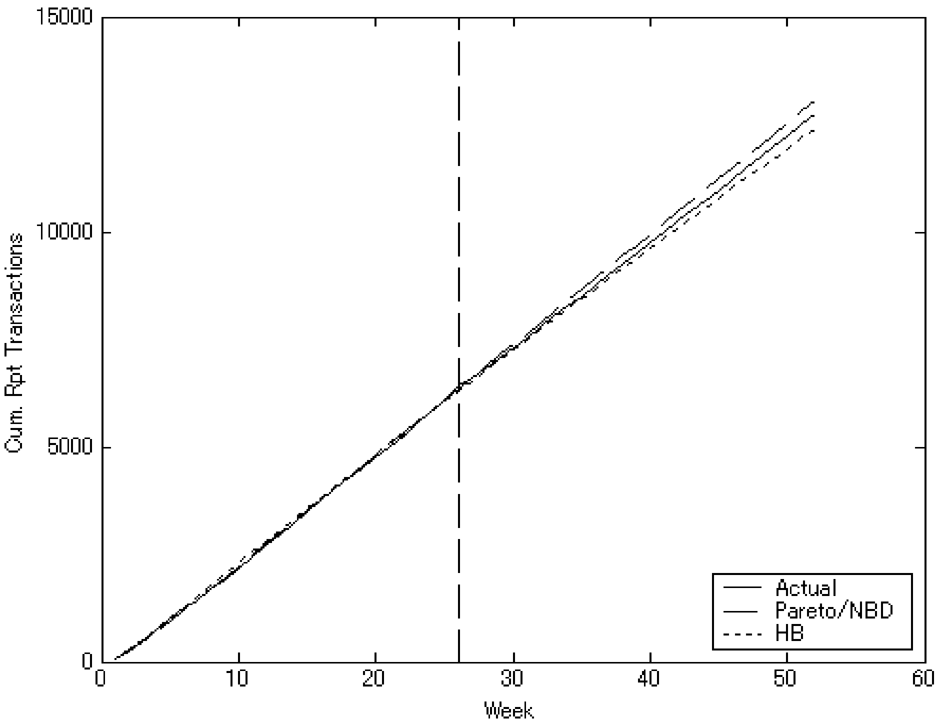


Figure 3. Conditional Expectation of Future Transactions

The proposed HB model shows a better forecast than Pareto/NBD model on the predicted number of transactions during the validation period, averaged across individuals and conditional on the number of purchases made during the calibration period.

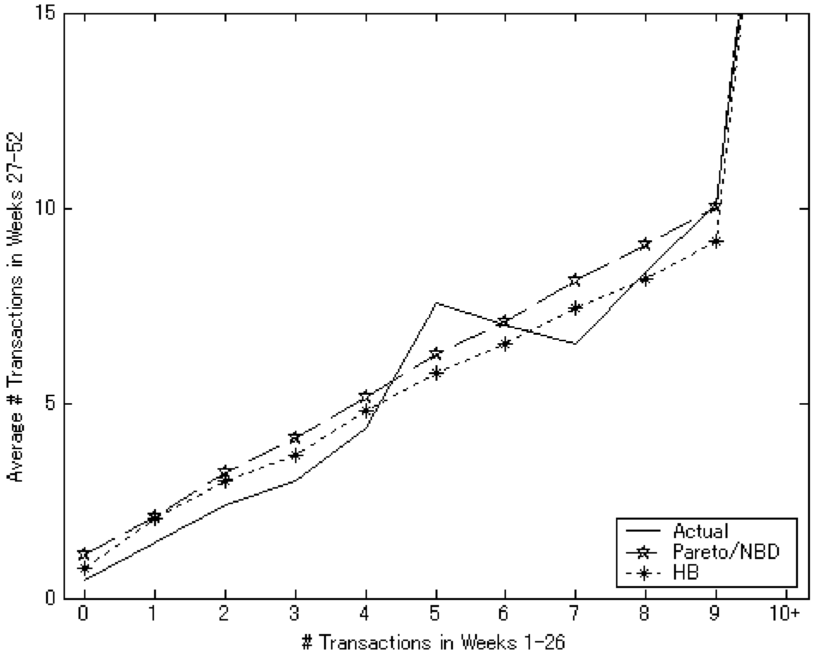


Figure 4. Conditional Expectation of Future Spending

The proposed HB model shows a better forecast than Pareto/NBD model on the predicted amount of spending during the validation period, averaged across individuals and conditional on the number of purchases made during the calibration period.

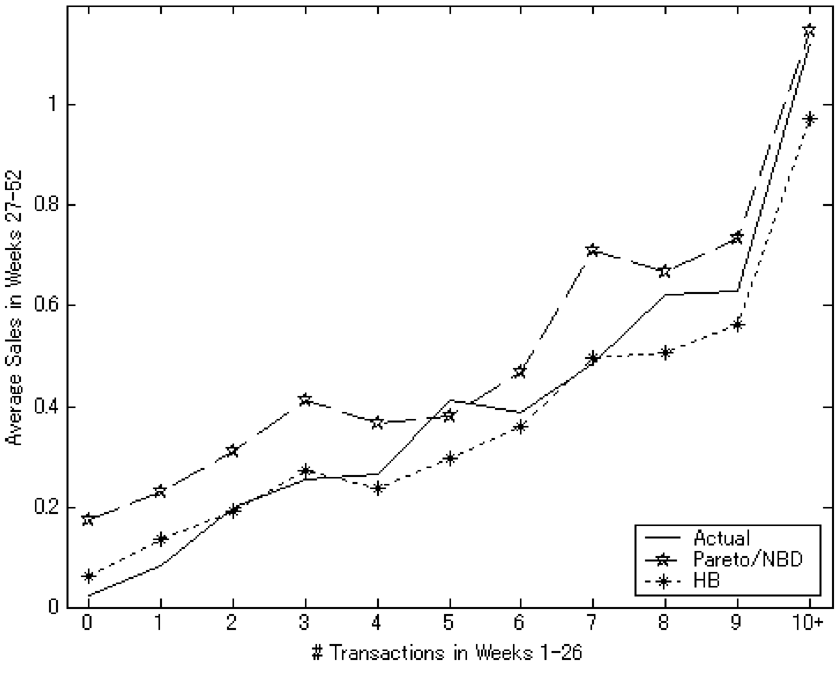


Figure 5. Scatter Plots of Posterior Means λ and η

Each point corresponds to one of the 400 households estimated. Note the negative correlation between purchase rate (lamda) and spending (eta).

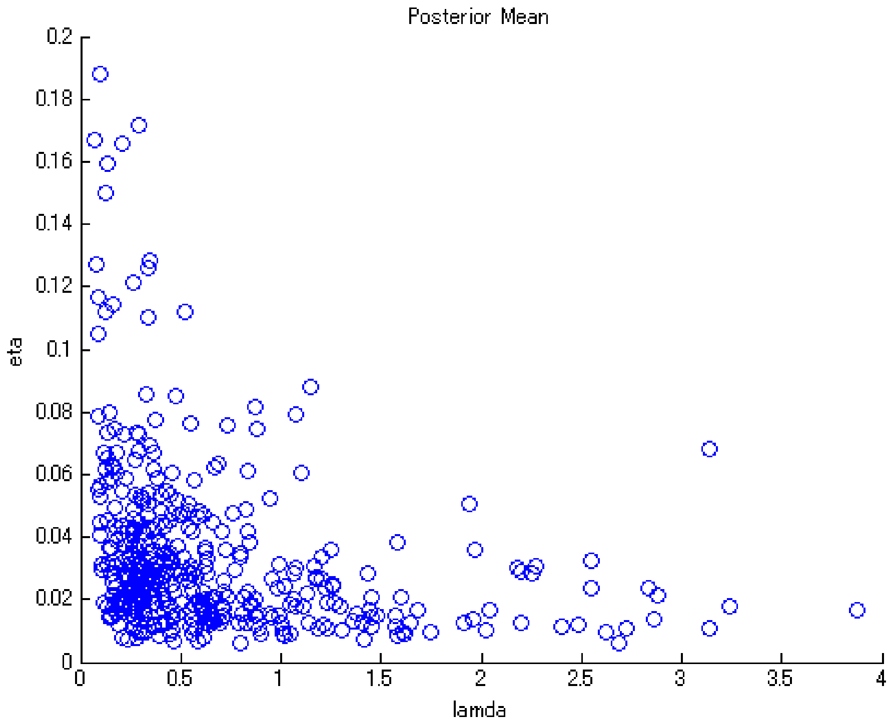


Figure 6. Gain Chart for CLV based on HB model and Simple Recency

RFM scoring, whereby three measures (RFM) are combined with equal weights, seems to provide a rather accurate ordering of CLV. However, customer ranking using of just one measure of recency is not that accurate.

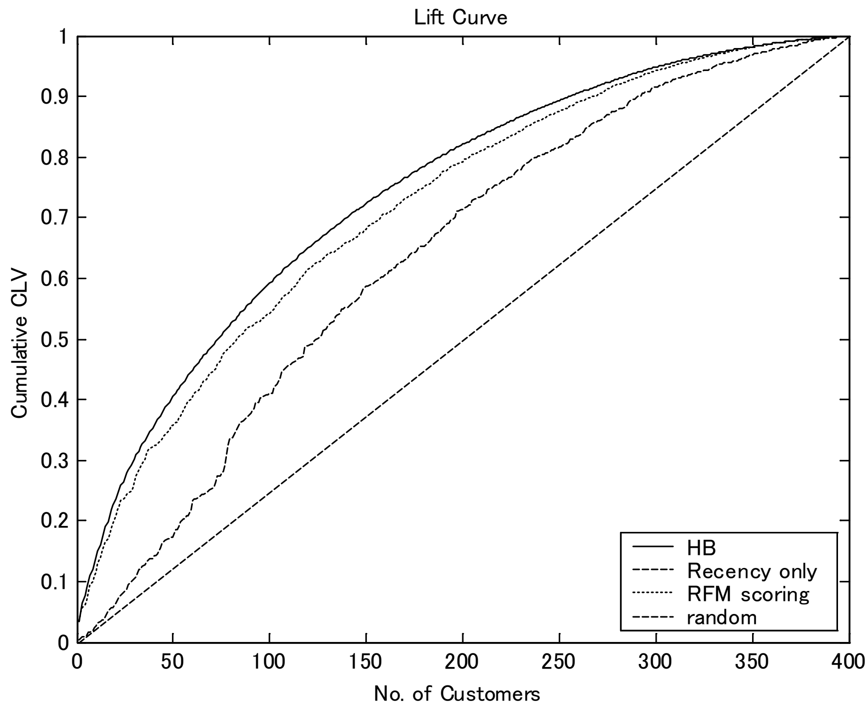


Figure 7. Retention Response Function

Investment level of *c* per customer would bring him/her to become active with probability *r(c)*. The function can be approximated by an exponential curve.

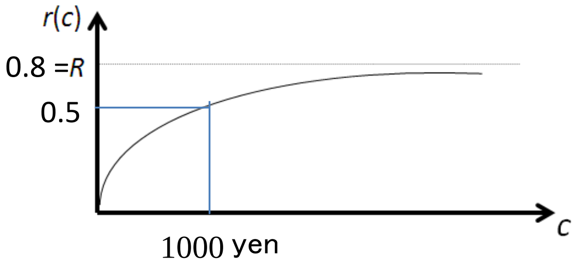
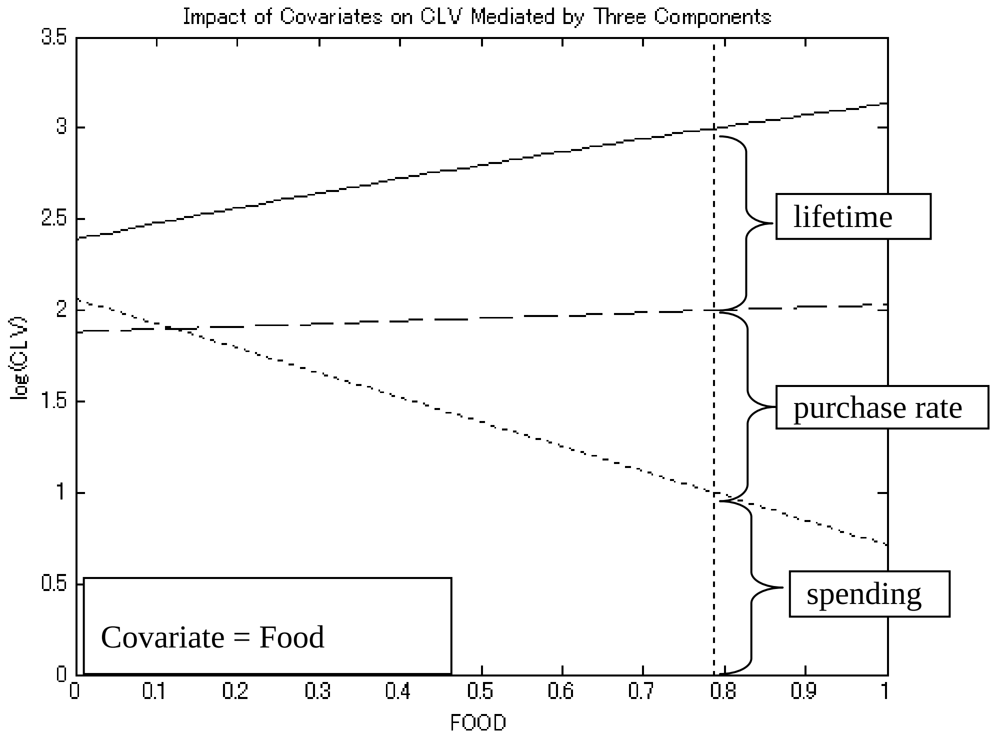
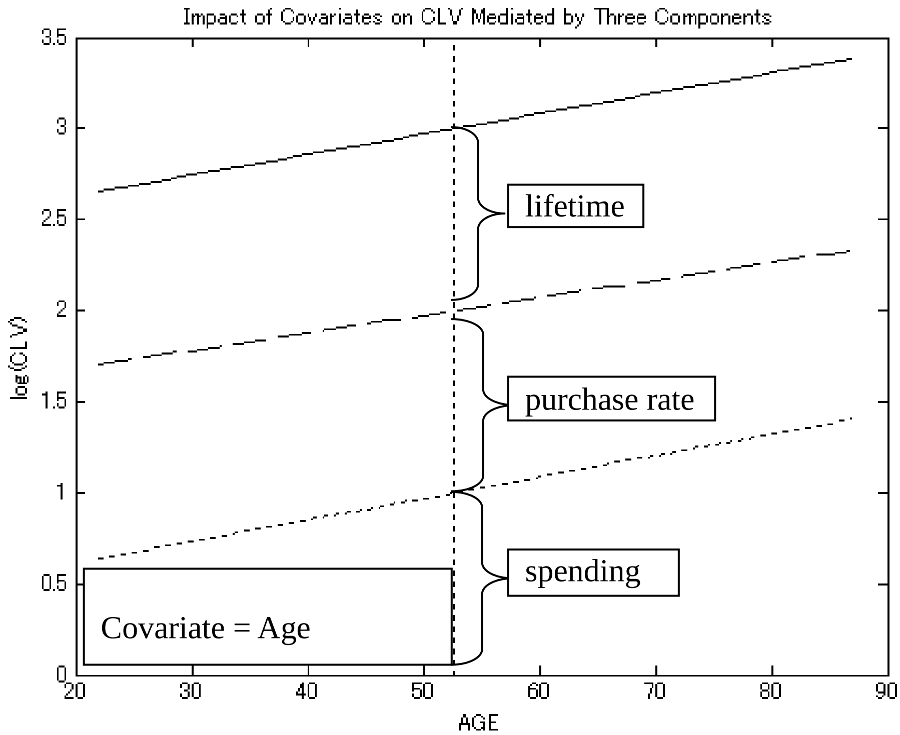


Figure 8. Impact of Covariates on CLV Decomposed into Three Components (Department store)

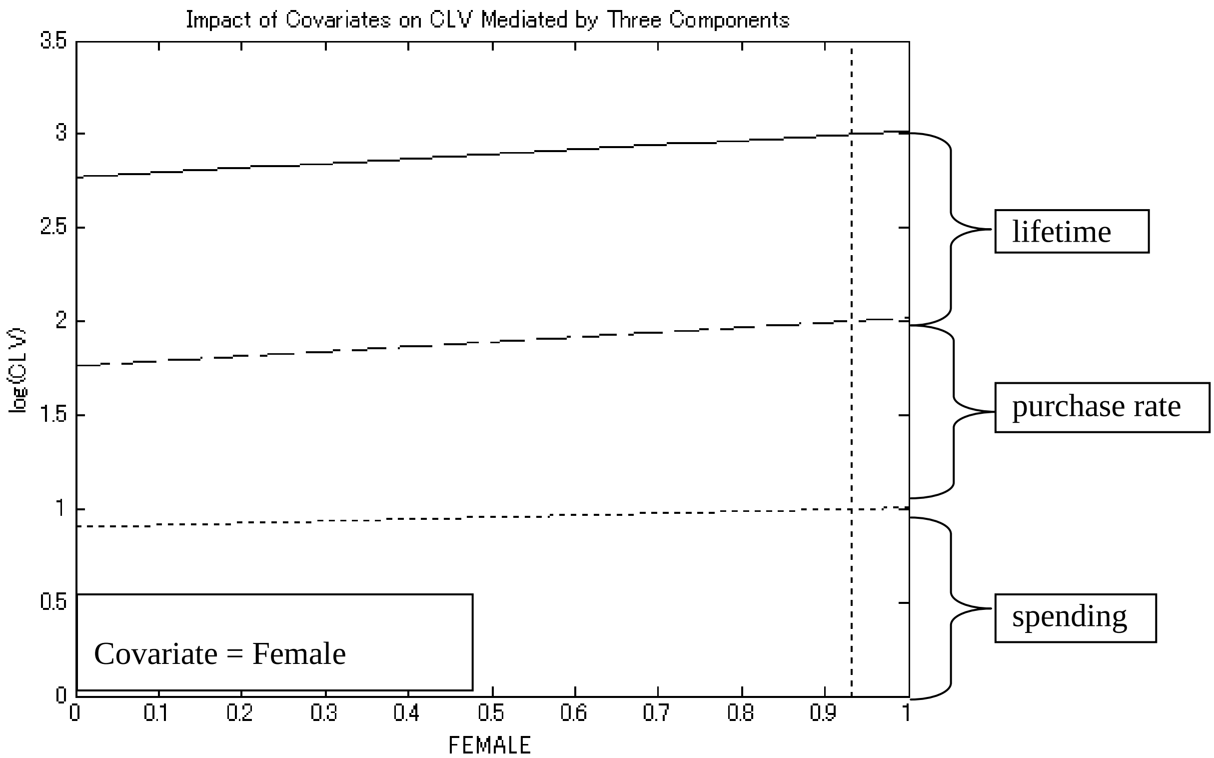
The dotted vertical line indicates the mean value of the covariate under consideration.



Increasing the fraction of food buyers improves lifetime and purchase rate, but decreases spending per purchase, yielding a net increase in the overall CLV.

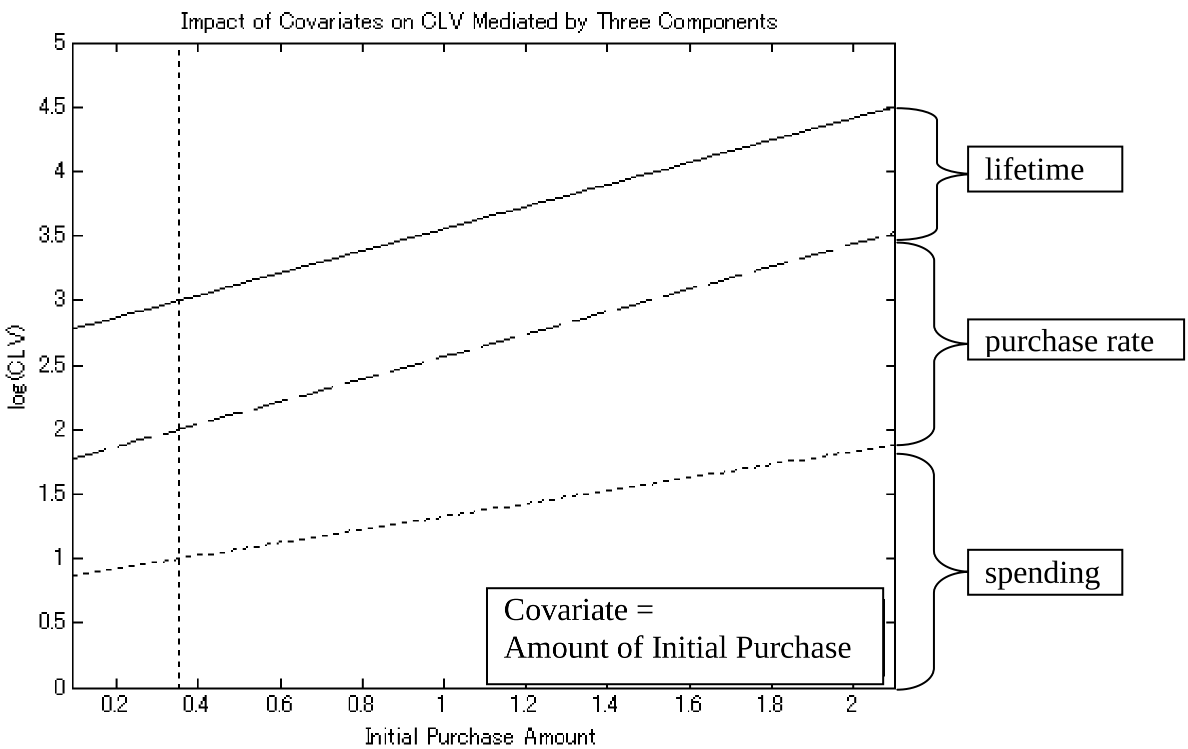


Increasing the fraction of elderly people increases the spending without much influence on lifetime and purchase rate, thereby resulting in a net increase in the overall CLV.

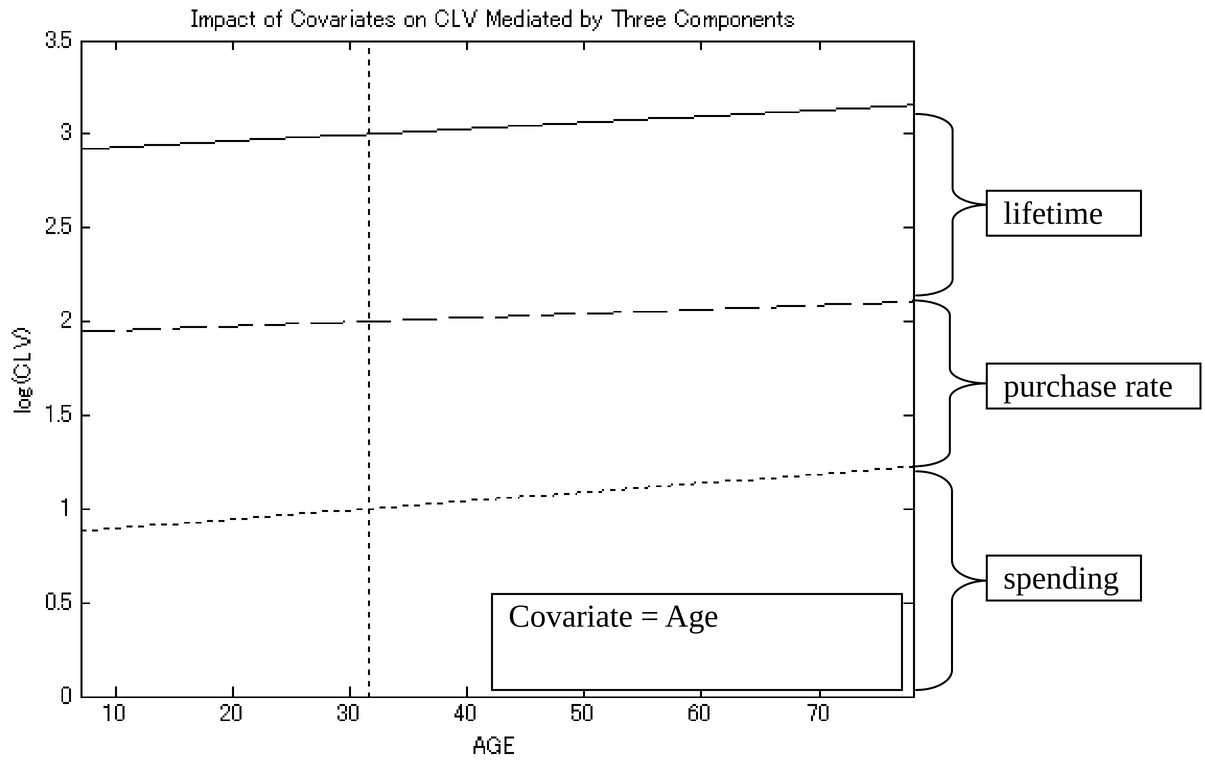


Increasing the fraction of females leads to little improvement in all three components and, hence, a negligible increase in the overall CLV.

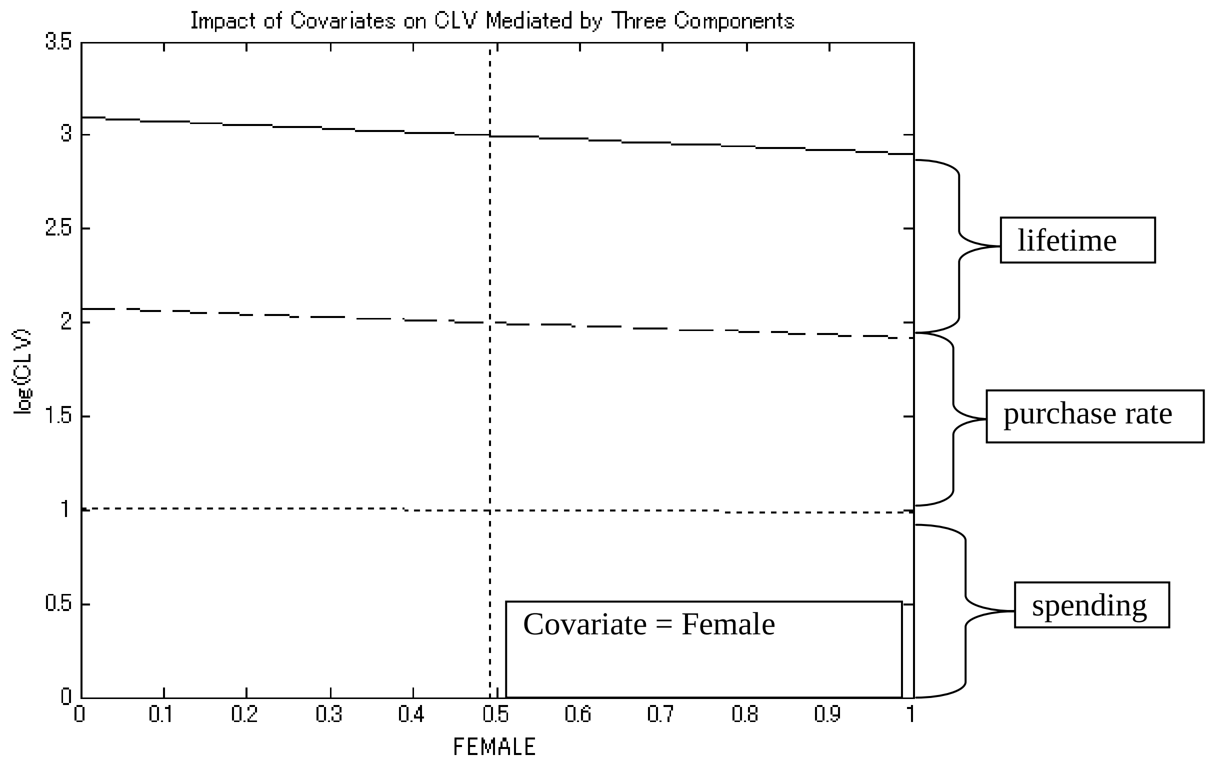
Figure 9. Impact of Covariates on CLV Decomposed into Three Components (CD chain)



A higher initial purchase amount is related to higher CLV by increasing the purchase rate and spending with almost no change in lifetime.



Older customers are associated with a lower purchase rate, longer lifetime, and higher spending per purchase with a positive net contribution to CLV.



Female customers are associated with a lower purchase rate, shorter lifetime, and less spending with a negative net contribution to CLV.