CIRJE-F-408

"Counting Your Customers" One by One: An Individual Level RF Analysis Based on Consumer Behavior Theory

Makoto Abe
University of Tokyo

March 2006

“COUNTING YOUR CUSTOMERS” ONE BY ONE:

消費者行動理論に基づいた個人レベルの **RF** 分析

阿部　誠
東京大学大学院経済学研究科

要約

CRM で重要な概念である顧客の生涯価値を計算するには、顧客の離脱率または維持率を把握することが必要である。しかし離脱する顧客は単に購買を止めるだけで、年会費などの支払い義務がないような“契約に基づかない状況”では、わざわざ離脱を申告することは稀だ。通常このような場合、企業は独自の経験則に基づいて、例えば顧客が 3 ヶ月購買しなければ離脱したと判断したりする。実務家の間でよく使われる RFM 分析では、(RECENCY = 3 ヶ月)のようなアドホックで一律なルールが基本になっているが、ここには２つの大きな問題がある。第 1 に、このルールが主観的なことである。なぜ 2 ヶ月や 4 ヶ月でなく、3 ヶ月なのだろうか？　２つ目の問題は、マーケティングの基本的概念である顧客の異質性を無視していることである。同じ 3 ヶ月の RECENCY でも、購買間隔が長い顧客は離脱の心配が無いが、購買間隔が短い顧客は離脱している可能性が高いであろう。つまり離脱率の推測に顧客の異質性に配慮する必要があるだろう。この問題は、Schmittlein et al. (1987)らが Pareto/NBD モデルを使った“counting your customers”フレームワークによって 20 年ほど前に研究したが、今日のマーケティングでは個々の顧客に焦点をあてた、よりミクロレベルの分析が求められている。

本論文では、Pareto/NBD モデルにおけるロバストな消費者行動の仮定（ポアソン購買プロセスとメモリレス離脱プロセス）は残しつつ、個人ごとにパラメータを推定することによって顧客の異質性をモデル化することを提案する。手法としては階層ベイズモデルを MCMC 法によって推定する。このモデルでは、Pareto/NBD モデルと違って 2 つの行動プロセルの独立性を仮定する必要がなく、かつ顧客ごとの生存期間や維持率など、従来得られなかった CRM に有用な指標が求められる。この研究では顧客の購買予測をベンチマークである Pareto/NBD モデルと比較する。スキャンパネルデータを使ったモデルの拡張では､RF データに顧客の購買行動データやデモグラフィック情報を加えることによって、ロイヤル顧客はより多くの金額を使うのか、またはより多くの利益を生むのか、などの CRM に重要な示唆が得られることを示した。

Key words: CRM, direct marketing, customer lifetime, Poisson process, Bayesian method

“COUNTING YOUR CUSTOMERS” ONE BY ONE:

AN INDIVIDUAL LEVEL RF ANALYSIS BASED ON

CONSUMER BEHAVIOR THEORY

Makoto Abe

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

March 20, 2006

"COUNTING YOUR CUSTOMERS" ONE BY ONE: AN INDIVIDUAL LEVEL RF ANALYSIS BASED ON CONSUMER BEHAVIOR THEORY

ABSTRACT

In customer relationship management (CRM), ad hoc rules are often employed to judge whether customers are active in a "non-contractual" setting. For example, a customer is considered to have dropped out if he or she has not made purchase for over three months. However, for customers with a long interpurchase time, this three-month time frame would not apply. Hence, when assessing customer attrition, it is important to account for customer heterogeneity. Although this issue was recognized by Schmittlein et al. (1987), who proposed the Pareto/NBD "counting your customers" framework almost 20 years ago, today's marketing demands a more individual level analysis.

This research presents a proposed model that captures customer heterogeneity through estimation of individual-specific parameters, while maintaining theoretically sound assumptions of individual behavior in a Pareto/NBD model (a Poisson purchase process and a memoryless dropout process). The model not only relaxes the assumption of independence of the two behavioral processes, it also provides useful outputs for CRM, such as a customer-specific lifetime and retention rate, which could not have been obtained otherwise. Its predictive performance is compared against the benchmark Pareto/NBD model. The model extension, as applied to scanner panel data, demonstrates that recency-frequency (RF) data, in conjunction with customer behavior and demographics, can provide important insights into direct marketing issues, such as whether long-life customers spend more and are more profitable.

Key words: CRM, direct marketing, customer lifetime, Poisson process, Bayesian method

1. INTRODUCTION

In CRM, it is important to know which customers are likely to be active and to be able to predict their future purchase patterns. This, in turn, allows the firm to take customized marketing action most suitable to each customer, as well as to estimate its current and future customer base for strategic planning. Under a “non-contractual” setting, however, consumers do not declare that they become inactive, but simply stop conducting business with the firm. To judge customer attrition, practitioners often use ad hoc rules, for instance, a customer is considered to have dropped out if he or she has not made a purchase for over three months.

There are two problems with this kind of judgment, however. First, it is not clear why the period of inactivity is three months rather than two or four months. Although the criterion of three months may be based on the experience of the firm, it hardly seems objective. Second, the criterion ignores customers’ differences in purchase frequency. Given the same period of nonpurchase, customers with a long interpurchase time may still be active, whereas those with a short interpurchase time are more likely to be inactive. As such, recognition of customer heterogeneity is a fundamental concept in marketing.

Using the framework of a BCG portfolio matrix, Figure 1 depicts the contribution of customers when one uses this type of ad hoc judgment in an RF analysis. Inactive customers (Problem Children and Dogs) are first isolated based on some cutoff in recency (e.g., three months), then active customers are further separated into the best (Stars) and the remaining (Cash Cows), using a frequency measure. Here, two criteria, recency and frequency, are considered independently. When recency and frequency are taken into account simultaneously, however, this interpretation changes, as seen in Figure 2.

Insert Figures 1 and 2 about here.

A subset of “Star” customers, who exhibits recencies that are longer than expected from their high frequency or purchase (i.e., interpurchase time), should be labeled as “Problem Children” (shaded upper-left triangle), requiring immediate attention before they become permanently inactive. Additionally, a subset of “Dog” customers, who exhibits recencies that are shorter than expected from their low frequencies, results in a surprising contribution to the firm, and hence these customers (shaded lower-right triangle) are labeled as “Cash Cows.” Failure to capitalize on these customer segments in the shaded triangles means a loss of opportunity for the firm.

This problem was first recognized by Schmittlein, Morrison, and Colombo (1987) (hereafter referred to as SMC). Based on common hypotheses about consumer behavior, SMC proposed a Pareto/NBD model that accounts for the relationship between recency and frequency and derived the probability of an individual customer being active at a particular point in time. In their model, consumer behavior is characterized by: (1) Poisson purchase (with purchase rate parameter λ) and (2) exponential lifetime (with dropout rate parameter μ). Further, λ and μ follow independent gamma distributions, which are formulated as a mixture distribution model. Although their work is highly regarded and follow-up research has been conducted (Fader, Hardie, and Lee, 2005a, 2005b; Reinartz and Kumar, 2000, 2003; Schmittlen and Peterson 1994), it is the increasing importance of new types of marketing, such as database marketing, CRM, and one-to-one marketing, that has brought this model to the attention of researchers and practitioners.

In this research, the behaviorally based RF analysis of SMC and others is extended to suit to the micro focus of today's marketing. While adopting the theoretically sound behavioral assumptions of SMC, the proposed approach captures customer heterogeneity through estimation of individual-specific parameters with a hierarchical Bayesian framework. In particular, this approach maintains the behavioral model of SMC, but: (1) replaces the analytical part of the heterogeneity mixture distribution with a simulation method and (2) incorporates unobservable measures such as a customer lifetime and an active/inactive binary indicator into the model as latent variables. By avoiding analytical aggregation, the approach leads to a simpler and cleaner model that provides eight advantages, as described below.

1. Conceptual simplicity. The analytical expression of the probability of being active and its complicated derivation, which SMC claim to be their main result (equations (11)-(13) and the appendix in their paper), can be skipped.

2. Estimation ease. Parameter estimation of the mixture distributions, which is investigated extensively in Schmittlein and Peterson (1994), also can be skipped.

3. Computational ease. Multiple evaluations of a non-standard Gauss hypergeometric function that is used in estimating a Pareto/NBD model are not necessary. To ease the computational burden, Fader, Hardie, and Lee (2005a) proposed a simplified BD/NBD model that closely approximates a Pareto/NBD model.

4. Model flexibility. The proposed model is more flexible in that the independence of purchase rate and dropout rate parameters, a crucial assumption in a Pareto/NBD model, need not hold. The parameter estimate of a Pareto/NBD model might be biased if this independence assumption were violated. The proposed model not only accommodates correlated data, but also allows the performing of statistical inference of the independence assumption on data, as described in (6) below.

5. Estimation of latent variables at the individual level. Purchase rate λ and dropout rate μ are estimated at the individual level. These parameters cannot be obtained from a Pareto/NBD model that is based on an empirical Bayes formulation, as will be explained in Section 3.1. A scatter plot of the posterior means of individual level λ and μ can be used to assess the independence assumption of a Pareto/NBD model. Other useful customer-specific statistics that could not have been obtained otherwise from a Pareto/NBD model include an expected lifetime and a 1-year retention rate.

6. Estimation of the correct measure of error. A Bayesian framework based on the MCMC simulation method used here does not produce a point estimate; rather, it produces a posterior distribution of parameters being estimated, providing a correct measure of error necessary for statistical inference. As will be shown in the subsequent empirical analysis, the distribution of the correlation between $\log(\lambda)$ and $\log(\mu)$ allows a formal testing of the independence assumption in a Pareto/NBD model.

7. Ease of model extension. Hierarchical models, whereby customer-specific parameters are a function of covariates, can be constructed and estimated with ease.

(a) Schmittlein and Peterson (1994) calibrate a Pareto/NBD model separately for each segment specified by the SIC code. The proposed model, by including segmentation variables in a hierarchical manner, allows estimation of all segments simultaneously, thereby increasing the degrees of freedom. The model also can incorporate non-nominal explanatory variables.

(b) To investigate the impact of customer characteristic variables on profitable lifetime duration, Reinartz and Kumar (2003) pursue a two-step approach: a lifetime duration is first estimated from RF data using a Pareto/NBD model, and then a proportional hazard model is constructed to link the lifetime duration (dependent variable) with characteristic variables (explanatory variables). A hierarchical model, whose dropout parameter is a function of customer characteristics, can be estimated in one step, providing the correct measures of error for statistical inference.

8. Exact Bayesian paradigm. The approach pursued by SMC is a so-called empirical Bayes, whereby the same data are used for the likelihood (customer specific purchase and survival functions) as well as for estimating the prior (mixture distribution), resulting in the overestimation of precision.[1] Although no threat is posed if the sample size is large or the mixture distribution is estimated from separate data, empirical Bayes is an approximation of a hierarchical Bayes method in the Bayesian paradigm (Gelman, Carlin, Stern, and Rubin 1995).

In the next section, the proposed model is described and compared against the NBD/Pareto model of SMC. Section 3 explains a simulation method for the estimation. Using data taken from a textbook by Franses and Paap (2002), Section 4 presents an empirical analysis, comparing the model's performance with that of the NBD/Pareto model. Section 5 contains a model extension, whereby purchase rate λ and dropout rate μ are linked to customer characteristic variables and potential marketing insights are sought. Section 6 presents the conclusions, limitations of the model, and future directions.

---

[1] The BD/NBD model, proposed by Fader, Hardie, and Lee (2005a), also uses an empirical Bayes method in their mixture distribution model, thereby suffering from similar complications.

2. PROPOSED MODEL VERSUS NBD/PARETO MODEL

2.1. Model Assumptions

This section provides an explanation of the assumptions of the proposed model.

Individual Customer

A1. Poisson purchases. While active, each customer makes purchases according to a Poisson process with rate λ.

A2. Exponential lifetime. Each customer remains active for a lifetime, which has an exponentially distributed duration with dropout rate μ.

These assumptions are identical to the behavioral assumptions of a Pareto/NBD model, and their validity has been studied by other researchers. Because their justification is documented elsewhere, including SMC, for brevity, further elaboration is not provided here.

Heterogeneity across Customers

A3. Individuals' purchase rates λ and dropout rates μ follow a multivariate lognormal distribution.

Unlike a Pareto/NBD model, whereby independent gamma distributions are assumed for λ and μ, this assumption permits a correlation between purchase and dropout processes. There are several reasons for the lognormal assumption.

(a) Bayesian updating of a multivariate normal (hence lognormal) is a standard procedure and easy to compute. The distribution can readily accommodate additional parameters through a hierarchical model, as will be shown in Section 5.

(b) The correlation between log(λ) and log(μ) can be obtained through the variance-covariance matrix of the lognormal distribution. A correlated bivariate distribution with gamma marginals is rather complicated (Park and Fader, 2004).

(c) In all of the previous studies using a Pareto/NBD model (Fader, Hardie, and Lee, 2005; Batislam, Denizel, and Filiztekin, 2004; Reinartz and Kumar, 2000, 2003; SMC, 1987; Schmittlein and Peterson, 1994), the shape parameter of the gamma distributed dropout rate μ (denoted as *s* in SMC) was estimated to be less that 1, implying that the expectation of active lifetime τ diverges to infinity (Equation (9) in SMC [1987]). Considering that customers eventually dropout (for various reasons, including natural causes such as death), a lognormal distribution seems more appropriate, at least for a prior.

The impact of the difference in the mixture distributions between a gamma and a lognormal will be evaluated in the subsequent empirical application.

2.2. Mathematical Notations

Figure 3 depicts the notations of SMC for recency and frequency data (x, t, T), which we will follow. The first transaction occurs at time 0 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 (x-th repeat) occurring at t. Hence, recency is defined as T-t. τ is an unobserved customer lifetime. Using mathematical notation, the previous model assumptions can be expressed as follow.

Insert Figure 3 about here.

(A1) $$P[x \mid \lambda, \tau > T] = \frac{(\lambda T)^{x}}{x!} e^{-\lambda T} \qquad x = 0,1,2,.$$

(A2) $$f(\tau) = \mu e^{-\mu \tau} \qquad \tau \geq 0$$

(A3) $$\begin{bmatrix} \log(\lambda) \\ \log(\mu) \end{bmatrix} \sim MVN\left(\theta_0 = \begin{bmatrix} \theta_\lambda \\ \theta_\mu \end{bmatrix}, \Gamma_0 = \begin{bmatrix} \sigma_\lambda^2 & \sigma_{\lambda\mu} \\ \sigma_{\mu\lambda} & \sigma_\mu^2 \end{bmatrix} \right)$$

where MVN denotes a multivariate normal distribution.

Some useful individual-level statistics were derived in the appendix. Similar derivations can be found in SMC (1987) and Fader, Hardie, and Lee (2005b).

3. ESTIMATION

3.1. Introducing Latent Variables

Our estimation approach is guided by taking into consideration the reason for not being able to estimate λ and μ individually in the empirical Bayes framework of a Pareto/NBD model.

In an empirical Bayes framework of a Pareto/NBD model:

Prior: λ_i ~ Gamma(r, α), μ_i ~ Gamma(s, β)

if active at T_i,

posterior: λ_i| $data_i$ ~ Gamma($r+x_i$, $\alpha+T_i$)

posterior: μ_i| $data_i$ ~ Gamma(s, $\beta+T_i$)

if inactive at T_i and dropout at $y_i<T_i$,

posterior: λ_i| $data_i$ ~ Gamma($r+x_i$, $\alpha+y_i$)

posterior: μ_i| $data_i$ ~ Gamma(s+1, $\beta+y_i$)

The above implies that the gamma distributions for λ and μ cannot be updated individually unless unobserved variables (i.e., whether customer i is active at T_i and, if not, the dropout time $y_i<T_i$) are known. Thus, let us introduce these unobservables as latent variables in our model.[2] For notational simplicity, subscript *i* is dropped in the following discussion. *z* is defined as 1 if a customer is active at time T and 0 otherwise. Another latent variable is a dropout time *y* when *z* = 0 (i.e., inactive). If we know *z* and *y*, then the likelihood function for RF data (*x, t, T*) becomes the following simple expression for *x*>0.[3]

Case z=1 (customer is active at T)

$$
\begin{aligned}
&P(x-\text{th purchase at } t \,\&\, \text{active until T} \,\&\, \text{no purchase between } [t,T]) \\
&= \frac{\lambda^x t^{x-1}}{\Gamma(x)} e^{-\lambda\, t} \times e^{-\mu T} \times e^{-\lambda(T-t)} \\
&= \frac{\lambda^x t^{x-1}}{\Gamma(x)} e^{-(\lambda+\mu)T}
\end{aligned}
$$

Case *z* = 0 (customer dropout at *y*≤*T*)

$$
\begin{aligned}
&P(x-\text{th purchase at } t \,\&\, \text{no purchase between } [t,y] \,\&\, \text{dropout at } y \le T) \\
&= \frac{\lambda^x t^{x-1}}{\Gamma(x)} e^{-\lambda\, t} \times e^{-\lambda(y-t)} \times \mu e^{-\mu y} \\
&= \frac{\lambda^x t^{x-1}}{\Gamma(x)} \mu e^{-(\lambda+\mu)y} \qquad (\text{t} \le y \le T)
\end{aligned}
$$

Combining the two cases, a more compact notation for the likelihood function can result.

(1) $$L(x,t,T \mid \lambda,\mu,z,y) = \frac{\lambda^x t^{x-1}}{\Gamma(x)} \mu^{1-z} e^{-(\lambda+\mu)\{zT+(1-z)y\}}$$

---

[2] Introduction of the latent variables is necessary because direct estimation of a Pareto/NBD model by standard software, such as WinBUGS, results in non-convergence due to the irregular shape of the likelihood function.

[3] If *x*=0, there is no repeat purchase and *t* = 0. Thus Γ(*x* = 0) and t^{x-1} are undefined. The appropriate likelihood function is $e^{-(\lambda+\mu)T}$ for $z=0$, and $\mu e^{-(\lambda+\mu)y}$ for $z=1$. Hence,

Because we observe neither z nor y, however, we treat them as missing data and apply a data augmentation technique (Tanner and Wong, 1987). To simulate z in our MCMC estimation procedure, we can use the following expression for the probability of a customer being active at T, or equivalently $z = 1$, derived in the appendix.

(2) $$P[\tau > T \mid \lambda, \mu, T, t] = P[z = 1 \mid \lambda, \mu, T, t] = \frac{1}{1 + \frac{\mu}{\lambda + \mu}\left[e^{(\lambda+\mu)(T-t)} - 1\right]}.$$

3.2. Estimation by Data Augmentation

Because parameter estimates for the purchase and dropout processes will be customer specific, index i (i=1,..,I) is reinstated to indicate individual customers. Let us denote the customer specific parameters as θ_i = [log(λ_i), log(μ_i)]', which is normally distributed with mean θ_0 and variance-covariance matrix Γ_0 as in (A3). Our objective is to estimate parameters $\{\theta_i, y_i, z_i, \forall i; \theta_0, \Gamma_0\}$ from observed recency and frequency data $\{x_i, t_i, T_i; \forall i\}$.

3.3. Prior Specification

To be consistent with the mixture distribution of λ_i and μ_i, the prior for λ_i and μ_i must be a lognormal as in (A3). The parameters of this lognormal, θ_0 and Γ_0 (i.e., hyper-parameters), are, in turn, estimated in a Bayesian manner with a multivariate normal prior and an inverse Wishart prior, respectively.

$$\theta_0 \sim MVN(\theta_{00}, \Sigma_{00}), \qquad \Gamma_0 \sim IW(\nu_{00}, \Gamma_{00})$$

Equation (1) becomes $L(x,t,T \mid \lambda, \mu, z, y) = \mu^{1-z} e^{-(\lambda+\mu)\{zT + (1-z)y\}}$.

These distributions are standard in a normal (and hence lognormal) model. Constants (θ_{00}, Σ_{00}, ν_{00}, Γ_{00}) are chosen to provide a very diffuse prior for the hyper-parameters θ_0 and Γ_0.

3.4. MCMC Procedure

We are now in a position to estimate parameters $\{\theta_i, \tau_i, z_i, \forall i; \theta_0, \Gamma_0\}$ using an MCMC method. To estimate the joint density, we sequentially generate each parameter, given the remaining parameters, from its conditional distribution until convergence is achieved. The procedure is described below.

[1] Set initial value for $\theta_i^{(0)}\ \forall i$.

[2] For each customer i,

[2a] generate $\{z_i \mid \theta_i\}$ according to equation (2).

[2b] If $z_i = 0$, generate $\{y_i \mid z_i, \theta_i\}$ using a truncated exponential distribution.

[2c] Generate $\{\theta_i \mid z_i, y_i\}$ using equation (1).

[3] Generate $\{\theta_0, \Gamma_0 \mid \theta_i; \forall i, \}$ using a standard normal update.

[4] Iterate [2]~[3] until convergence is achieved.

Below are explanations for each step.

[2a] θ_i obtained from the previous iteration is exponentiated to transform to λ_i and μ_i, which, in turn, can be plugged into equation (2) to compute $P(z_i = 1)$.

[2b] $z_i = 0$ means customer i dropped out after the last purchase before T_i. Thus, y_i must follow the exponential distribution (A2) with $\mu = \mu_i$ and truncation such that $t_i < y_i < T_i$.

[2c] Given z_i and y_i, equation (1) is used (through multiplication by the prior) to generate λ_i and μ_i, which are then transformed to θ_i by taking their logarithm. Because these distributions are not in a standard form, an independent Metropolis-Hasting algorithm (Allenby and Rossi, 2005) is used to generate λ_i first and then μ_i, where the proposal distribution is chosen to be lognormal.

4. EMPIRICAL ANALYSIS

We now apply the proposed model (hereafter denoted as the HB model [hierarchical Bayes]) to real data and make a comparison with the Pareto/NBD model. A dataset is taken from a textbook by Franses and Paap (2001, p. 25), downloadable from their website. These A.C. Nielsen scanner panel data from Sioux Falls, South Dakota contain interpurchase times of liquid laundry detergents for 400 customers over 106 weeks during the late 1980s. The distribution of the interpurchase times (Figure 2.9 in their book) resembles an exponential distribution in the aggregate, supporting our assumption of a Poisson purchase process.

Like most scanner panel data, the data are left-censored. That is, the database does not contain purchase records prior to July 1986, dating back to the initial purchase of each household. Accordingly, customer lifetime must be interpreted with care, conditional on being active in July 1986. Because the dataset provides only interpurchase times, but not the exact dates of purchases, all households are assumed to have made their first purchase (trial) at the same time. Dates for the remaining purchases (repeats) are computed from their interpurchase times. The first 53 weeks of the data are used for model calibration and the remaining 53 weeks are used for model validation. The number of repeat purchases during the calibration ranges from 0 to 49, with the average being 4.0, as seen in Figure 4.

Insert Figure 4 about here.

The MCMC steps were put through 15,000 iterations, of which the last 5,000 were used to construct the posterior distribution of parameters. The convergence was monitored visually and checked with the Geweke test (Geweke, 1992). The dispersion of the proposal distribution in the Metropolis-Hasting algorithm was chosen such that the acceptance rate remained at about 40% to allow even drawing from the probability space (Gelman, Carlin, Stern, and Rubin, 1995).

The parameters for the Pareto/NBD model were estimated by MLE to be $r = 2.15$, $\alpha = 25.88$, $s = 0.16$, $\beta = 30.20$, following the notations of SMC. The proposed HB model was compared against the benchmark Pareto/NBD model for fit in the calibration period and prediction in the validation period. For disaggregate performance measures, correlation and a mean squared error (MSE) between predicted and observed number of purchases for individual customers were used. For an aggregate measure, a root mean squared (RMS) fractional error between predicted and observed weekly cumulative transactions was used. Table 1 present the results for the Pareto/NBD and HB models.

Insert Table 1 about here.

Both models perform similarly and neither of them dominates. The HB model has some advantage in the calibration sample, but the difference is minor. The simple four-parameter Pareto/NBD exhibited a surprisingly robust performance, whereas the HB model could not capitalize on its much higher degrees of freedom. This is because only three data points (x_i, t_i, T_i) were used to estimate individual-specific parameters.

For a visual check of the model performance at the aggregate level, Figure 5 presents a weekly time-series tracking of the cumulative numbers of purchases for the two models, along with the actual number. The vertical dotted line at week 53 separates the calibration from the validation period. Except towards the end, both HB and Pareto/NBD models fit well with the observed data, with negligible differences. The actual sales level off towards the end, because observations for all households were assumed to have started at week 0, thereby advancing the timing of the actual purchases by household-specific amounts. Neither the HB nor the Pareto/NBD model is expected to reproduce this artifact caused by the way the dataset was constructed.

Insert Figure 5 about here.

For a visual check at the disaggregate level, Figure 6 shows the predicted number of transactions during the validation period, averaged across individuals and conditional on the number of transactions made during the calibration period. The use of this measure is suggested by Fader, Hardie, and Lee (2005a). Consistent with the time-series tracking, both models tend to overestimate during the validation period. Unusually high transactions (about 1.6) for 0 transactions in weeks 1-53 are caused by the way the sample is collected. Because the original dataset by Franses and Paap contains records of households who made at least one repeat purchase during the entire 106 weeks, those households who do not make any repeat purchases during the calibration period automatically make purchases in the validation, causing as artificial “kink.”

Insert Figure 6 about here.

Figure 7 is a scatter plot of the posterior means of λ and μ for the 400 households in the dataset. One can clearly see a high degree of heterogeneity in the purchase and dropout rates. Two households who purchased 49 and 27 times during the calibration period (see Figure 4) are seen as outliers, with high values of λ. The L-shaped distribution is expected from a theoretical consideration. The estimate of lifetime (i.e., dropout rate μ) is greatly influenced by the timing of the last purchase. If the last purchase occurs early, a shorter lifetime is estimated, and vice versa. Under low frequency, for example with a single repeat purchase, the last purchase could occur anytime during the observation period, causing a high variation in the lifetime estimate. When frequency is high, unless all purchases are clustered at the beginning, the last purchase tends to occur towards the end. This results in a longer lifetime (smaller μ).

Insert Figure 7 about here.

Figure 8 presents a histogram of the correlation between log(λ) and log(μ), which is obtained from 5,000 MCMC draws for the variance-covariance matrix (hyper-parameter) of the lognormal mixture distribution. Its mean is -0.176 and the 2.5 and 97.5 percentiles are -0.470 and 0.103, respectively. For this database, therefore, the independence assumption between purchase and dropout processes appears to hold. SMC does not offer specific means to test the independence assumption of λ and μ.

Insert Figure 8 about here.

Table 2 presents six customer-specific statistics for randomly chosen 20 customers: posterior means of λ_i and μ_i, an expected lifetime, a retention rate after one year, the probability of being active at the end of the calibration period, and an expected number of transactions during the validation period.[4] The first four statistics could not have been obtained from the Pareto/NBD model, but are quite useful in CRM. For example, more accurate evaluation of a customer lifetime value would be possible with a customer-specific retention rate. Statistics in the last two columns are claimed by SMC to be the main result, with complicated expressions (equations (11)-(13) and (22) in their paper). With the HB model, the simple individual level formulas (4) and (5) in the appendix are applied to each draw of λ and μ from the MCMC procedure, and their means are calculated.

Insert Table 2 about here.

----

[4] Because $E[f(\mu)] \neq f(E[\mu])$, $f(\mu)$ is calculated for each simulation draw of μ, all of which are then averaged across draws to obtain these statistics, for an expected lifetime, a median instead of a mean is reported.

In sum, the HB model: (1) imposes fewer assumptions (the correlation between λ_i and μ_i is permitted), accommodating a wider range of data, and (2) provides parameter estimates at customer level (posterior means of λ_i and μ_i, an expected lifetime, a retention rate) that can be useful in CRM, and (3) predicts transactions as well as the Pareto/NBD model.

However, the real advantage of the HB model is its conceptual simplicity. Because modeling effort can be completed strictly at the individual behavior level, one can extend the model without having to deal with the complex and sensitive operation of aggregation over heterogeneous customers (i.e., mixture distribution).

5. MODEL EXTENSION

This section illustrates a model extension, whereby transaction rate λ and dropout rate μ are linked to customer characteristic variables. Such a model can offer insights into the profile of customers with long lifetime and frequent purchases. If the characteristics are demographic variables, the model allows a manager to pursue acquisition of prospective customers whose behavioral (transaction) data are not yet collected.

5.1. Model and Estimation

A straightforward approach is to specify the logarithm of λ_i and μ_i with a linear regression as follows.

(A3') $$\begin{bmatrix} \log(\lambda_i) \\ \log(\mu_i) \end{bmatrix} \equiv \theta_i = \beta' d_i + e \qquad \text{where } e \sim MVN(0, \Gamma_0)$$

d_i is a K×1 column vector that contains K characteristics of customer i. β is a K×2 parameter vector and e is a 2×1 error vector that is normally distributed with mean 0 and variance Γ_0. This formulation replaces θ_0 in the previous section with $\beta' d_i$. When d_i contains only a single element of 1 (i.e., an intercept only), this model reduces to the previous no covariate case. The MCMC step is modified accordingly (replacing θ_0 by $\beta' d_i$) and the third step is changed to

[3’] $\{\beta, \Gamma_0 \mid \theta_i, \forall i\}$ using a standard multivariate normal regression update.

See Bayesian textbooks elsewhere for details on the multivariate normal regression update (Congdon, 2001; Gelman, Carlin, Stern, and Rubin, 1995; Rossi, Allenby, and McCulloch, 2005).

5.2. Empirical Analysis

The detergent dataset used in the previous section (Franses and Paap, 2001, p. 25) also contains some behavioral and demographic information on these households. Three customer characteristic variables are constructed from the raw data. The first is the average dollar spent, in hundreds of dollars, per shopping trip for a household. It is constructed by summing all the detergent and non-detergent expenses of a household and dividing by the number of the household’s purchase occasions.[5] The second is the deal proneness of a household. It is defined as the fraction of detergent purchases bought on deal (feature, display, or both). The third variable is household size.

Table 3 presents the result of three HB models with increasing complexity, along with that of the previous Pareto/NBD model. The no-covariate HB model is the one used in Section 4. With respect to both aggregate and disaggregate measures, all HB models

[5] Because the dataset provides the volume of the last, but not the current, purchase occasion, the detergent and non-detergent expenditures are shifted by one purchase occasion. The error should have a minimal impact because we are aggregating across all purchase occasions.

perform as well as Pareto/NBD does in calibration and validation. The reported coefficients are posterior means, and the 2.5 and 97.5 percentiles in the parentheses provide their standard error-like measures. Note that the left hand side of the regression is a logarithm of λ and μ, and thus the magnitude of the intercept must be interpreted accordingly.

Insert Table 3 about here.

As covariates are added, estimated coefficients remain stable. The exceptions are average spending and household size, whose correlation is moderately high at 0.362. We will focus on M2, which has the highest marginal likelihood. For $\log(\lambda)$, the significant covariates at the 5% level are deal proneness and household size. This indicates that customers who buy detergent on deal tend to buy less frequently, perhaps doing so only when detergent is promoted. Larger households tend to purchase detergent more frequently, which is intuitive. For $\log(\mu)$, household size is significant at the 5% level, indicating that larger households have a shorter lifetime. One possible explanation is that they switch stores more often, seeking lower prices.

No significant relationship was found between detergent purchase frequency and average spending. Additionally, the dropout rate was not related to average spending or deal proneness. This means that long-life customers do not necessary spend more. A positive relationship between the dropout rate and deal proneness would have suggested that long-life customers (i.e., have a low dropout rate) are more profitable (i.e., are less deal prone). We did not find such evidence from this database.

6. CONCLUSIONS

6.1. Summary

A great deal has changed since the work of SMC almost 20 years ago. Advances in information technology, combined with conceptual development in database marketing, CRM and one-to-one marketing, allow even unsophisticated firms to pursue customized marketing actions of some form at the individual customer level. Marketing has seen some shift from an aggregate to a disaggregate focus. In keeping with this, an individual level RF analysis, based on consumer behavior theory was developed, resulting in an HB model, which was then estimated by a MCMC method.

The HB model presumes three tried and true assumptions of SMC: (1) a Poisson purchase process, (2) a memoryless dropout process (i.e., constant hazard rate), and (3) heterogeneity across customers, while relaxing SMC's independence assumption of the purchase and dropout processes. Because customer heterogeneity is captured as a prior in a hierarchical Bayesian framework, instead of through a mixture distribution, the entire modeling effort can bypass all the complications associated with aggregation, which is left to MCMC simulation. The advantages include: (1) conceptual simplicity, (2) estimation ease, (3) computational ease, (4) model flexibility, (5) estimation of latent variables, (6) estimation of correct error measures, (7) ease of model extension, and (8) the exact Bayesian paradigm.

The HB model was shown to perform well in the empirical analysis using publicly available data. Outputs included individual level λ_i and μ_i, an expected lifetime, a retention rate, the probability of being active, and an expected number of future transactions, of which the first four were not available from a Pareto/NBD model.

The conceptual simplicity of the HB model has lead to an estimable model, in which λ and μ are a function of customer characteristic variables. The model extension applied to scanner panel data demonstrates that RF data, in conjunction with customer behavior and demographics, can provide important insights into direct marketing issues such as whether long-life customers spend more and are more profitable.

The current study also confirmed the sound performance of a Pareto/NBD model, which predicted transactions as well as the HB model, if not better. It appears that the infinite expected customer lifetime, caused by fitting a gamma distribution to the dropout rate, is not a problem. A Pareto/NBD model should continue to perform well, as long as the independence of the purchase and dropout processes holds. Here, the HB model can provide useful information to assess the validity of this assumption through: (1) a scatter plot of the posterior means of individual level λ and μ and (2) a distribution of the correlation between $\log(\lambda)$ and $\log(\mu)$.

6.2. Limitations and Suggestions for Future Research

One weakness of the HB model is that the closed form expressions on the statistics for a "randomly" chosen customer, such as the probability of being active and the expected number of future purchases, do not exist. Closed form can provide intuitive understanding of the aggregate market behavior as a whole by calculating comparative statistics. In the HB model, aggregate statistics must be constructed by simulation. Given that both Pareto/NBD and HB models have resulted in similar predictive performance, the two models can complement each other. A Pareto/NBD model can describe the aggregate customer response in a parsimonious manner for firms' strategic purposes, whereas the individual focus of the HB model could be used in actual operationalization of one-to-one marketing.

Several directions are possible in extending this research. One is a substantive investigation of the relationship between customer lifetime and profitability in non-contractual businesses. The current study is more methodological in nature and falls short of drawing any substantive conclusions on these issues. Relying on publicly available data, the detergent database was used only for illustrative purpose to demonstrate the potential of the HB model.

Pioneering research by Reinartz and Kumer (2000, 2003) can be improved upon in various ways using the HB model. First, the independence assumption of λ and μ in a Pareto/NBD model, on which their entire analysis was based, can be relaxed. Second, Reinartz and Kumer (2000) defined lifetime as the duration for which the probability of a customer being alive dropped below a threshold of c, after carefully justifying the value to be $c = 0.5$. That is still subjective, however. The estimate of individual μ available from the HB model can be used as an objective measure of customer lifetime. Third, the HB model can reveal the link between customer lifetime and characteristics in a one-step estimation, with accurate statistical inference, instead of the two-step estimation they employed.

The second natural direction is to extend the model from transaction to dollar amount by incorporating monetary value from RFM data. Such a model could provide valuable insights into customer lifetime value and customer equity, as was done by Fader, Hardie, and Lee (2005b) and Reinarts and Kumer (2000, 2003).

The third direction is to relax the assumption of the Poisson purchase process so that interpurchase time can take a more general form in distribution (Allenby, Leone, and Jen, 1999). A Poisson process implies random purchase occurrence with an exponentially distributed interpurchase time. While non-patrons might make purchases at random, loyal customers generally purchase at more regular intervals. A model that can capture behavioral differences in repeat purchase patterns beyond frequency could provide valuable insights into CRM. However, this extension puts more burdens on the part of data collection, because the analysis requires not just recency but all purchase timing.

It is the conceptual simplicity of the HB model that produces myriad possibilities for extension.

APPENDIX: Derivation of Survival Probability and Likelihood Function

Using Bayes rule, the survival probability can be derived from purchase history as follows.

$$
\begin{aligned}
P(\tau > T \mid \lambda, \mu, x, t, T) &= P(alive \mid history) \\
&= \frac{P(alive \,\&\, history)}{P(history)} \qquad (3) \\
&= \frac{P(history \mid alive)\, P(alive)}{P(alive \,\&\, history) + P(dead \,\&\, history)}
\end{aligned}
$$

Because the survival time is exponentially distributed, *P(alive)* is

$$P(alive) = P(\tau > T) = e^{-\mu T} .$$

Furthermore, the following two equations can be derived.

$$
\begin{aligned}
P(history \mid alive) &= P(x - \text{th purchase at } t \,\&\, \text{no purchase between } [t, T]) \\
&= \frac{\lambda^x t^{x-1}}{\Gamma(x)} e^{-\lambda t} \times e^{-\lambda(T-t)} \\
&= \frac{\lambda^x t^{x-1}}{\Gamma(x)} e^{-\lambda T}
\end{aligned}
$$

$$
\begin{aligned}
P(history \,\&\, dead) &= \int_t^T P(x - \text{th purchase at } t \,\&\, \text{no purchase between } [t, y] \,\&\, \text{die at } y \in [t, T])dy \\
&= \int_t^T \frac{\lambda^x t^{x-1}}{\Gamma(x)} e^{-\lambda t} \times e^{-\lambda(y-t)} \times \mu e^{-\mu y} dy \\
&= \frac{\lambda^x t^{x-1}}{\Gamma(x)} \mu \int_t^T e^{-(\lambda+\mu)\mathrm{y}} dy \\
&= \frac{\lambda^x t^{x-1}}{\Gamma(x)} \frac{\mu}{\lambda + \mu} \left\{ e^{-(\lambda+\mu)\mathrm{t}} - e^{-(\lambda+\mu)\mathrm{T}} \right\}
\end{aligned}
$$

Substituting the three equations above into Equation (3) leads to the survival probability formula.

$$P\left(\tau > T \mid \lambda, \mu, x, t, T\right) = \frac{\frac{\lambda^x t^{x-1}}{\Gamma(x)} e^{-\lambda T} \times e^{-\mu T}}{\frac{\lambda^x t^{x-1}}{\Gamma(x)} e^{-\lambda T} \times e^{-\mu T} + \frac{\lambda^x t^{x-1}}{\Gamma(x)} \frac{\mu}{\lambda + \mu} \left\{ e^{-(\lambda+\mu)t} - e^{-(\lambda+\mu)T} \right\}}$$

$$= \frac{1}{1 + \frac{\mu}{\lambda + \mu} \left\{ e^{(\lambda+\mu)(T-t)} - 1 \right\}} \qquad (4)$$

The expected number of transactions in the time period of t conditional on λ and μ can be derived as

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

Formulas for other relevant individual statistics are

$$\text{The expected lifetime} = \frac{1}{\mu}$$

$$\text{The retention rate after one year} = \exp(-52\mu) \qquad \text{where time unit is expressed in weeks}$$

Figure 1. Traditional RF Analysis

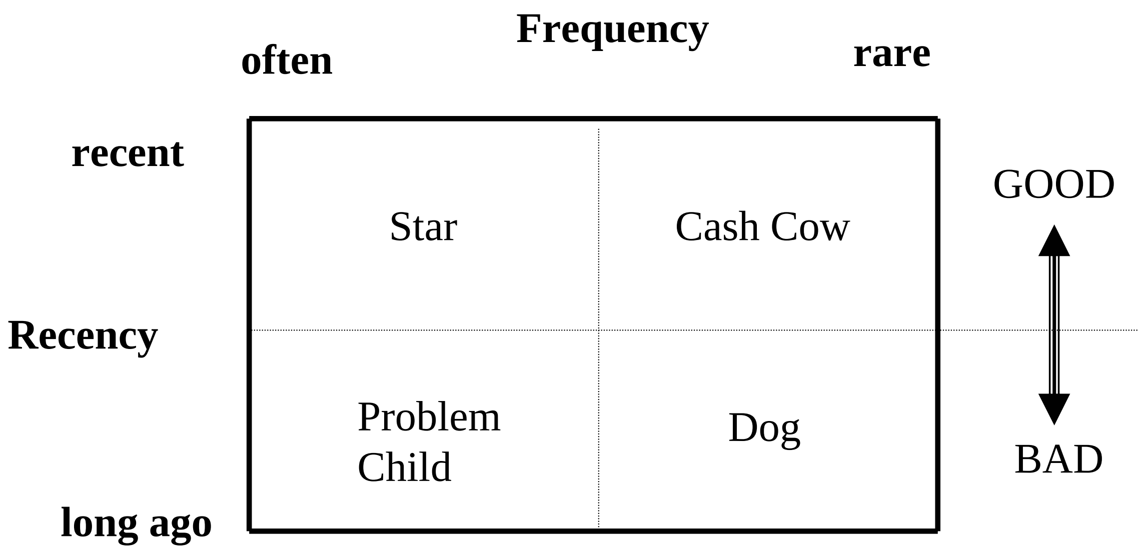

Figure 2. Proposed RF Analysis

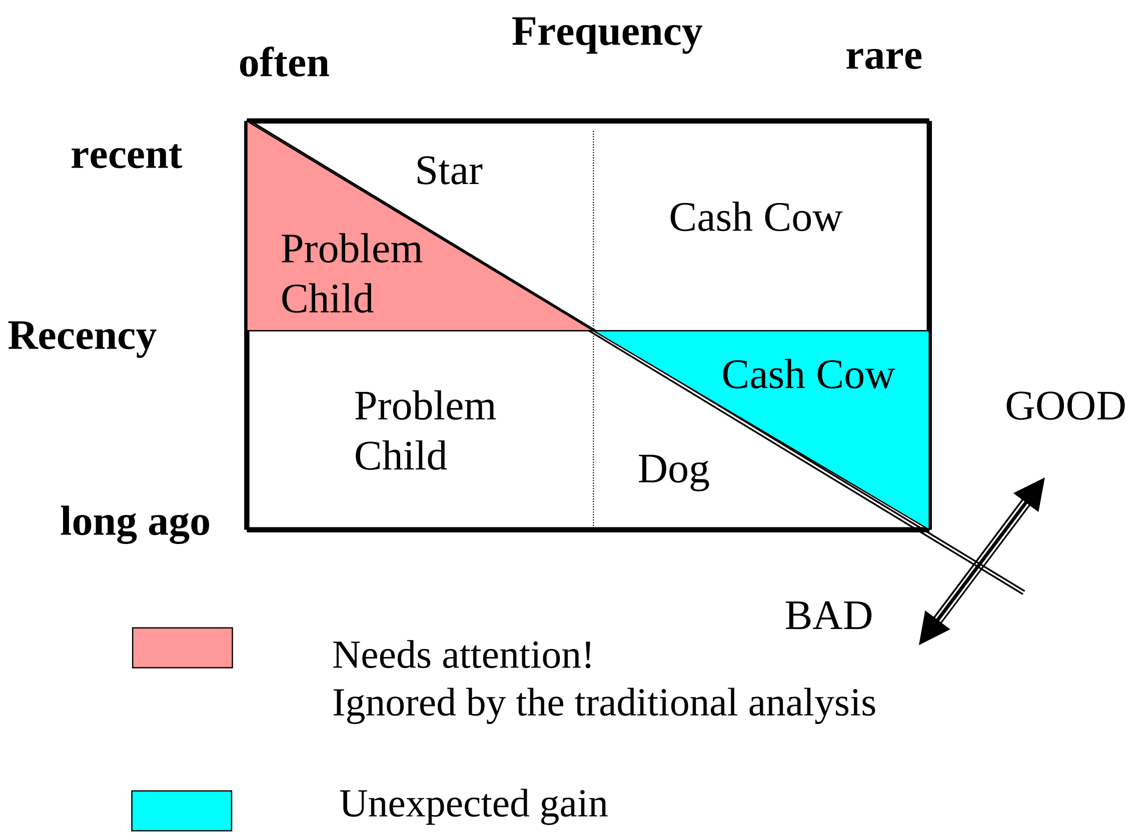

Figure 3. Notations for RF Data

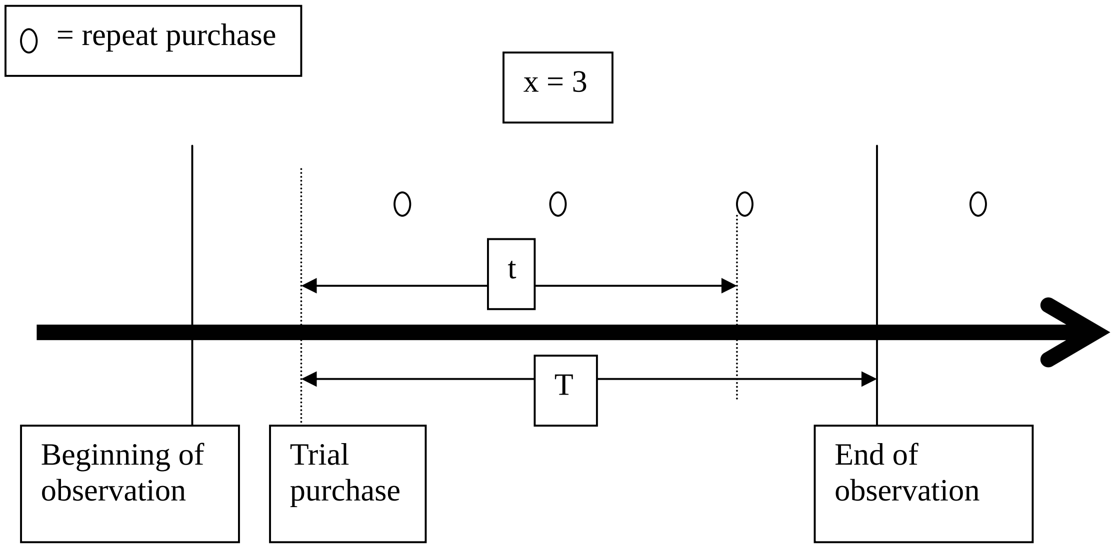

Figure 4. Histogram for the Number of Repeat Purchases

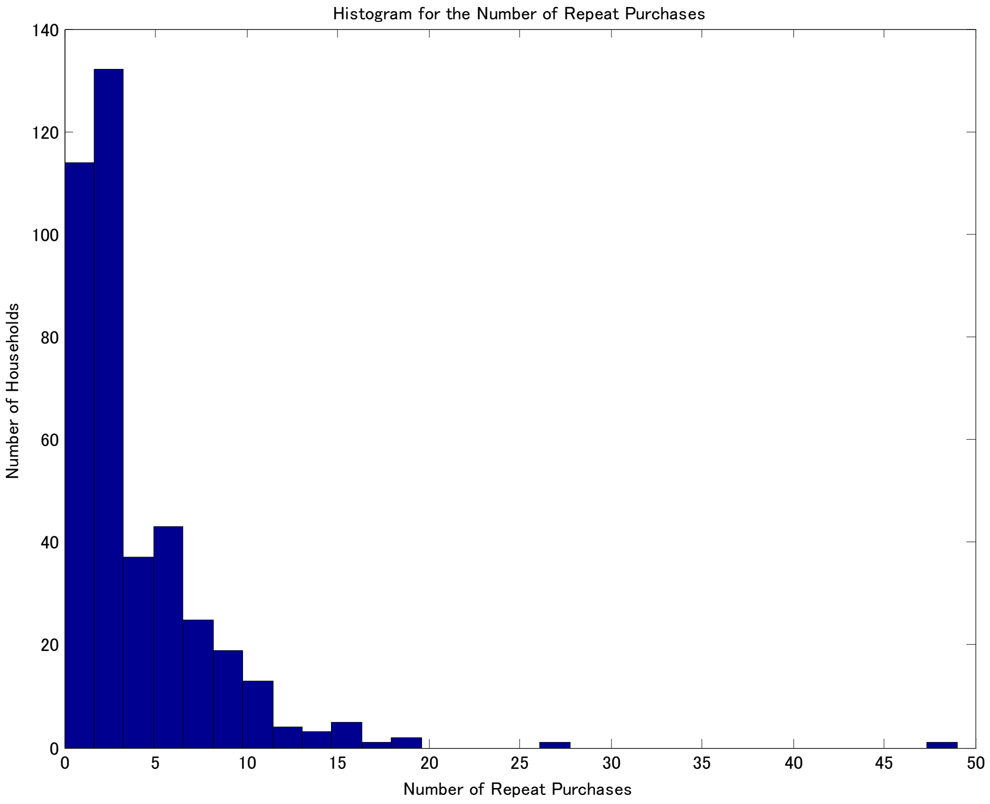

Figure 5. Weekly Time-series Tracking Plot

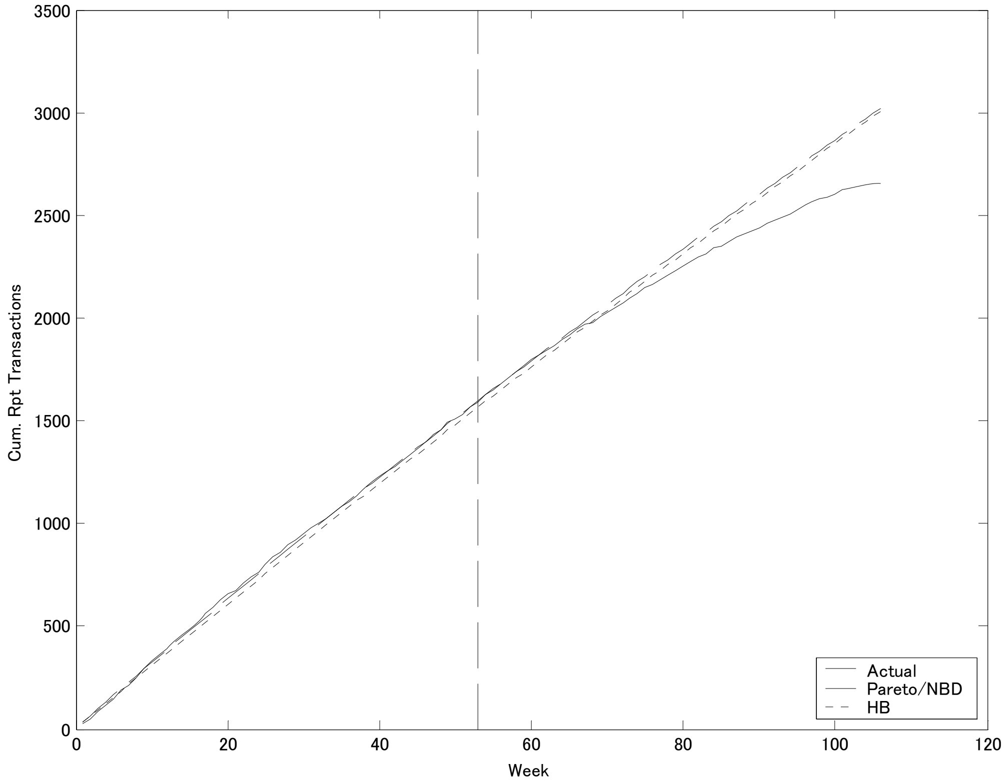

Figure 6. Conditional Expectation of Future Transactions

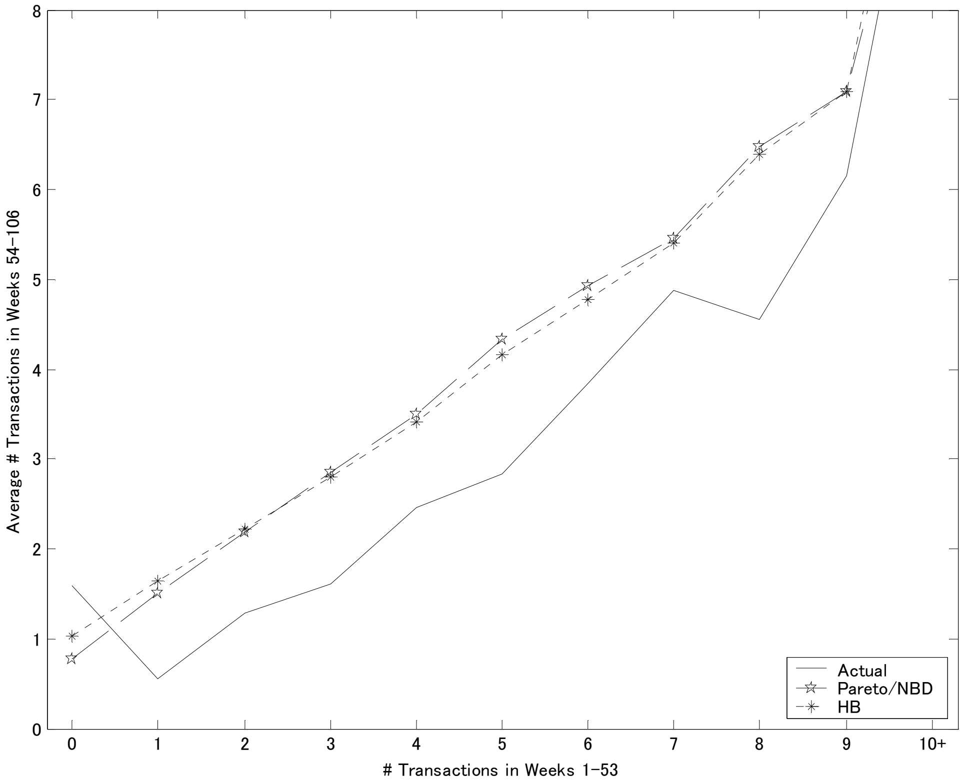

Figure 7. Scatter Plot of Posterior Means λ and μ

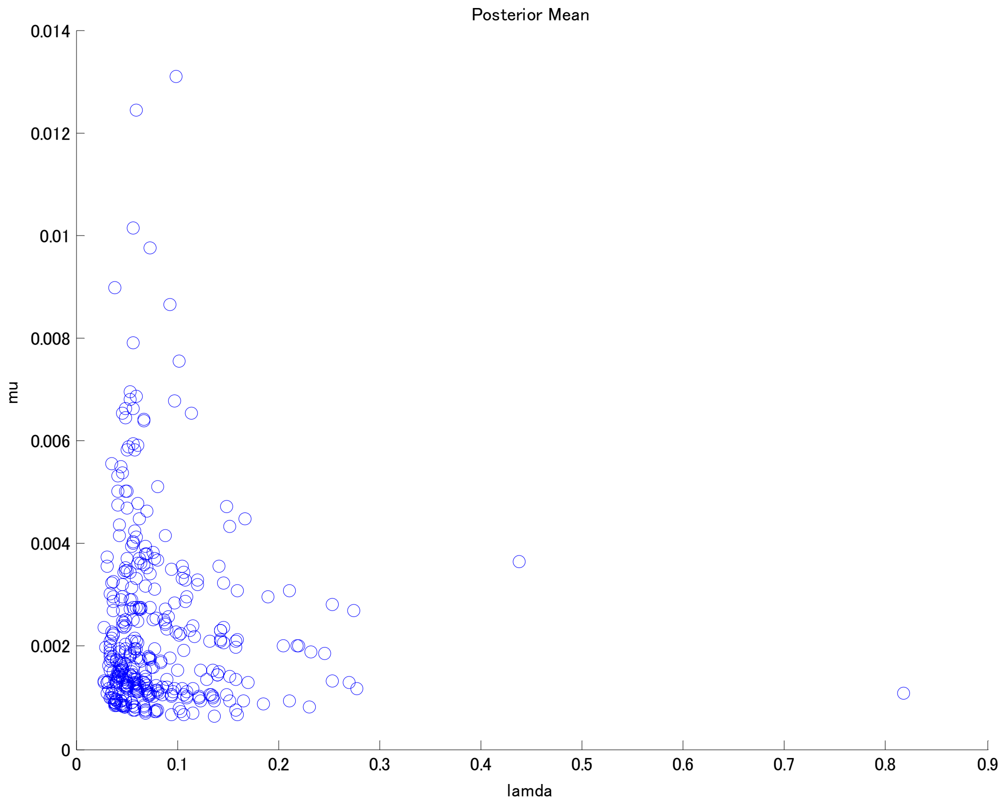

Figure 8. Distribution of the Correlation between log(λ) and log(μ)

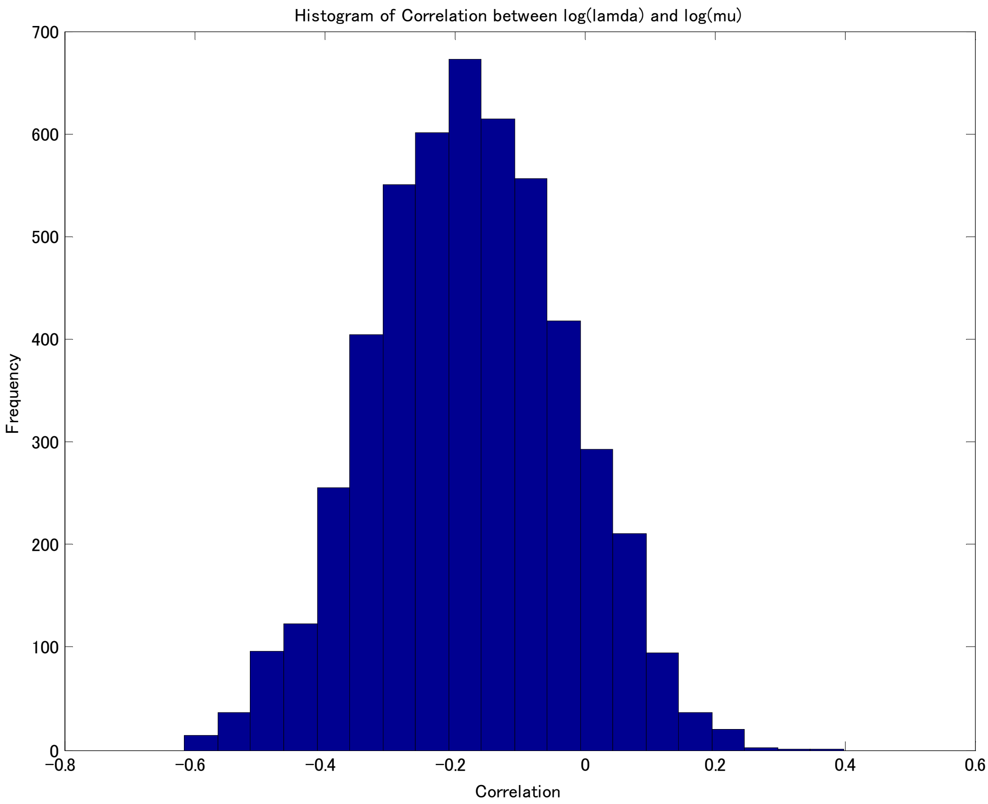

Table 1. Model Performance Result

Criterion		Pareto/NBD	HB model
Disaggregate Measure			
Validation	**correlation**	0.755	0.762
	MSE	7.988	7.951
Calibration	**correlation**	1.000	0.991
	MSE	2.312	1.486
Aggregate Measure			
Timeseries RMS	**Validation**	0.0604	0.0553
	Calibration	0.0867	0.0768
	Pooled	0.0747	0.0669

Table 2. Customer-Specific Statistics for Randomly Chosen 20 Customers

Customer	λ	μ	Median Expected lifetime (years)	1 year Retention rate	Probability of being active at the end of calibration	Expected number of future transactions
1	0.056122	0.0015447	19.725	0.927	0.92088	2.6194
2	0.054988	0.003055	9.9895	0.8659	0.96322	2.6109
3	0.039026	0.00081933	34.136	0.95923	0.98162	1.9899
4	0.070112	0.0007174	37.309	0.96407	0.99327	3.6262
5	0.040274	0.0098448	3.3742	0.68868	0.45462	0.68135
6	0.04938	0.0013984	20.901	0.93329	0.90381	2.2418
7	0.070327	0.001205	23.608	0.94121	0.9778	3.5355
8	0.11594	0.000717	40.704	0.9643	0.99935	6.0293
9	0.057639	0.0068565	4.5185	0.74647	0.44505	0.93857
10	0.059592	0.0059657	5.4833	0.77289	0.48776	1.098
11	0.14252	0.002307	12.476	0.89393	0.96573	6.892
12	0.068959	0.00071623	39.732	0.96434	0.9912	3.5588
13	0.15743	0.002052	13.076	0.90357	0.99222	7.8707
14	0.036983	0.00099033	30.126	0.95127	0.97413	1.8627
15	0.085826	0.0026205	10.272	0.87951	0.99807	4.2538
16	0.06875	0.0035991	7.6689	0.84285	0.99629	3.3306
17	0.058333	0.0019631	13.908	0.90709	1	2.945
18	0.057428	0.00078257	37.143	0.96125	0.99168	2.9619
19	0.059256	0.00078438	36.095	0.96096	0.95945	2.9336
20	0.084212	0.001103	26.337	0.94605	0.99025	4.3039

Table 3. Model Performance Result

(Figures in parentheses indicate the 2.5 and 97.5 percentiles)

		Pareto/NBD model	no-covariate HB model	M1 HB model	M2 HB model
Purchase rate λ	**Intercept**	---	-2.802 (-2.903, 2.705)	-2.790 (-2.961, -2.626)	-3.046 (-3.266, -2.834)
	Deal Proneness	---	---	-0.351 * (-0.621, -0.067)	-0.345 * (-0.615, -0.069)
	Ave. Spending	---	---	0.177 (-0.078, 0.435)	0.063 (-0.205, 0.335)
	Household Size	---	---	---	0.110 * (0.043, 0.179)
Dropout Rate μ	**Intercept**	---	-7.268 (-9.318,-5.975)	-6.926 (-9.006, -5.696)	-7.934 (-8.598, -7.264)
	Deal Proneness	---	---	-0.165 (-1.738, 1.249)	0.375 (-0.468, 1.254)
	Ave. Spending	---	---	-1.266 (-3.509, 0.398)	-0.056 (-0.946, 0.712)
	Household Size	---	---	---	0.397 * (0.145, 0.623)
correlation(log(λ), log(μ))		---	-0.176 (-0.470, 0.103)	-0.162 (-0.457, 0.148)	-0.196 (-0.562, 0.250)
log marginal likelihood		---	-2556.4	-2548.7	-2520.6
Disaggregate Measure					
Validation	**correlation**	0.755	0.762	0.760	0.758
	MSE	7.988	7.951	8.065	8.015
Calibration	**correlation**	1.000	0.991	0.990	0.985
	MSE	2.312	1.486	1.472	1.586
Aggregate Measure					
Timeseries RMS	**Validation**	0.0604	0.0553	0.0567	0.0537
	Calibration	0.0867	0.0768	0.0769	0.0760
	Pooled	0.0747	0.0669	0.0676	0.0658

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

REFERENCES

Allenby, Greg, Robert P. Leone, and Lichung Jen (1999), "A Dynamic Model of Purchase Timing with Application to Direct Marketing," *Journal of the American Statistical Association*, 94 (446), Applications and Case Studies, 365-374.

Batislam, Emine, Meltem Denizel, and Alpay Filiztekin (2004), "Lifetime Duration and Value of Grocery Retailer Customers," presentation at 2004 Marketing Science Conference, Rotterdam, Netherland.

Congdon, Peter (2001), *Bayesian Statistical Modelling*, London, UK: Wiley.

Franses, Philip Hans and Richard Paap (2001), *Quantitative Models in Marketing Research*, Cambridge, UK: Cambridge University Press.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin (1995), *Bayesian Data Analysis*, Boca Raton, Florida: Chapman & Hall.

Geweke, J. (1992), "Evaluating the Accuracy of Sampling-Based approaches to the Calculation of Posterior Moments," in J. M. Bernardo, J. M. Berger, A. P. Dawid and A. F. M. Smith, (eds.), *Bayesian Statistics 4*, 169-193, Oxford: Oxford University Press.

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2005a), "'Counting Your Customers' the Easy Way: An Alternative to the Pareto/NBD Model," *Marketing Science*, 24 (2), 275-284.

-----, ----- and ----- (2005b), "RFM and CLV: Using Iso-Value Curves for Customer Base Analysis," *Journal of Marketing Research*, 42 (4), 415-430.

Park, Young-Hoon and Peter S. Fader (2004), "Modeling Browsing Behavior at Multiple Websites," *Marketing Science*, 23 (3), 280-303.

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.

Rossi, Peter E., Greg Allenby and Rob McCulloch (2005), *Bayesian Statistics and Marketing*, London, UK: Wiley.

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.

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

Tanner, Martin A. and Wing Hung Wong (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82 (398), Theory and Methods, 528-540.