CIRJE-F-537

"COUNTING YOUR CUSTOMERS" ONE BY ONE: A Hierarchical Bayes Extension to the Pareto/NBD Model

Makoto Abe
University of Tokyo

January 2008; Revised in February 2009

“COUNTING YOUR CUSTOMERS” ONE BY ONE:

A Hierarchical Bayes Extension to the Pareto/NBD Model

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

January, 2008

Revised: February 17, 2009

* The manuscript has benefited greatly from comments and suggestions by Siddharth S. Singh, Sharad Borle, and Dipak C. Jain.

“COUNTING YOUR CUSTOMERS” ONE BY ONE:
A Hierarchical Bayes Extension to the Pareto/NBD Model

ABSTRACT

This research extends a Pareto/NBD model of customer-base analysis using a hierarchical Bayesian (HB) framework to suit today’s customized marketing. The proposed HB model presumes three tried and tested assumptions of Pareto/NBD models: (1) a Poisson purchase process, (2) a memoryless dropout process (i.e., constant hazard rate), and (3) heterogeneity across customers, while relaxing the independence assumption of the purchase and dropout rates and incorporating customer characteristics as covariates. The model also provides useful output for CRM, such as a customer-specific lifetime and survival rate, as by-products of the MCMC estimation.

Using three different types of databases --- music CD for e-commerce, FSP data for a department store and a music CD chain, the HB model is compared against the benchmark Pareto/NBD model. The study demonstrates that recency-frequency data, in conjunction with customer behavior and characteristics, can provide important insights into direct marketing issues, such as the demographic profile of best customers and whether long-life customers spend more.

Key words: CRM, direct marketing, customer lifetime, Bayesian method, MCMC

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 (Rust and Chung 2006, Sun 2006). 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/she has not made a purchase for over three months.

There are two problems with this kind of judgment. 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 chosen according to 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. Recognition of customer heterogeneity is critical in such a context.

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 recency-frequency (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 Pareto/NBD, 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 model and its estimation becomes simpler, thereby, permitting to accommodate various model extensions as follow.

First, 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, can be relaxed. 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.

Second, because the distribution of the purchase rate λ and dropout rate μ are estimated at the individual level as a byproduct of the MCMC method, the distribution of any customer-specific statistics that are functions of λ and μ can be obtained by simple algebra. Such statistics of managerial relevance include a probability of being active at a certain point in time, an expected lifetime, a 1-year survival rate, and an expected number of transactions in a future period. The fact that a distribution rather than a point estimate of a statistic is obtained also permits the application of statistical inference at the individual level without being restricted by the asymptotic properties.

Third, 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 achieve these in one step, providing the correct measures of error for statistical inference.

Mathematically, the approach pursued in a Pareto/NBD model is 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 (a mixture distribution), resulting in the overestimation of precision. 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 Pareto/NBD model. Section 3 explains the estimation by an MCMC method. Section 4 presents empirical analyses with three datasets of various types, comparing the model's performance against that of the Pareto/NBD model. Section 5 presents the conclusions, limitations of the model, and future directions.

2. PROPOSED MODEL VERSUS PARETO/NBD MODEL

2.1. Model Assumptions

This section provides an explanation for 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, further elaboration is not provided here for brevity.

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 2.3.

(b) Correlation between $\log(\lambda)$ and $\log(\mu)$ can be obtained through the variance-covariance matrix of the normal mixture distribution. A correlated bivariate distribution with gamma marginals is rather complicated (Park and Fader 2004).

The impact of the difference in the mixture distributions between a gamma and a lognormal would be the difference in fit between the Pareto/NBD and HB models, and that will be evaluated in the subsequent empirical analyses.[1]

2.2. Mathematical Notations

Figure 1 depicts the notations of SMC for recency and frequency data (x, t_x, T), which we will follow here. 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

[1] A gamma distribution is thought to be more flexible than a lognormal distribution because the former can have a mode at 0 or an interior mode, depending on whether the value of the shape parameter is less than 1 or not, whereas the latter accommodates only an interior mode. The author appreciates this insight provided by one of the reviewers.

the number of repeat transactions observed in the time period (0, *T*], with the last purchase (x-th repeat) occurring at t_x. Hence, recency is defined as T-t_x. τ is an unobserved customer lifetime. Using these mathematical notations, the preceding model assumptions can be expressed as follow.

< Insert Figure 1 about here >

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

where MVN denotes a multivariate normal distribution.

Some useful individual-level statistics that will be shown in the subsequent empirical analyses are derived in the online appendix. Similar derivations can be found in SMC and Fader, Hardie, and Lee (2005b).

2.3. Incorporating Covariates

A model that links purchase and dropout rates λ and μ to customer characteristics can offer insights into the profile of customers with frequent transactions and long lifetime. 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.

A straightforward approach is to specify the logarithm of λ_i and μ_i with a linear regression as follows, where index *i* is added to emphasize that the rate parameters are for customer *i*.

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

d_i is a K×1 column vector that contains K characteristics of customer *i*. β is a K×2 parameter vector and e_i 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.

3. ESTIMATION

3.1. Introducing Latent Variables

Our estimation approach is guided by exploiting the reason for not being able to estimate λ and μ individually in the empirical Bayes framework of a Pareto/NBD model. In the Pareto/NBD model, as shown in the online appendix:

Prior: $\lambda_i \sim gamma(r, \alpha)$, $\mu_i \sim gamma(s, \beta)$

if active at T_i,

posterior: $\lambda_i |\ data_i \sim gamma(r+x_i, \alpha+T_i)$

posterior: $\mu_i |\ data_i \sim gamma(s, \beta+T_i)$

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

posterior: $\lambda_i |\ data_i \sim gamma(r+x_i, \alpha+y_i)$

posterior: $\mu_i |\ data_i \sim gamma(s+1, \beta+y_i)$

The above implies that, 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, one cannot take advantage of the simple Bayesian updating for the conjugate priors. This is the very reason for the complex estimation process associated with the Pareto/NBD model. Thus, let us introduce these unobservables as latent variables in our model. 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_x, T) becomes the following simple expression for $x>0$.

Case z=1 (customer is active at T)

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

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

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

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

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

For x=0, there is no repeat purchase and t_x = 0. Thus $\Gamma(x = 0)$ and t_x^{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, Equation (5) becomes $L(x = 0, t_x, T \mid \lambda, \mu, z, y) = \mu^{1-z} e^{-(\lambda+\mu)\{zT+(1-z)y\}}$.

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 online appendix.

(6)
$$P[\tau > T \mid \lambda, \mu, T, t_x] = P[z = 1 \mid \lambda, \mu, T, t_x] = \frac{1}{1 + \frac{\mu}{\lambda+\mu}\left[e^{(\lambda+\mu)(T-t_x)} - 1\right]}.$$

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 $\beta' d_i$ and variance-covariance matrix Γ_0 as in equation (4). Our objective is to estimate parameters $\{\theta_i, y_i, z_i, \forall i; \beta, \Gamma_0\}$ from observed recency and frequency data $\{x_i, t_{x(i)}, T_i; \forall i\}$.

3.3. Prior Specification

Following equation (4), the prior for λ_i and μ_i is chosen to be lognormal. The parameters of this lognormal, β 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.

$$\beta \sim MVN(\beta_0, \Sigma_0), \qquad \Gamma_0 \sim IW(\nu_{00}, \Gamma_{00})$$

These distributions are standard in a normal (and hence lognormal) model. Constants (β_0, Σ_0, ν_{00}, Γ_{00}) are chosen to provide a very diffuse prior for the hyper-parameters β and Γ_0.

3.4. MCMC Procedure

We are now in a position to estimate parameters $\{\theta_i, y_i, z_i, \forall i; \beta, \Gamma_0\}$ 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 (6).

[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 (5).

[3] $\{\beta, \Gamma_0 \mid \theta_i, \forall i\}$ using a standard multivariate normal regression 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 (see equation (4)), which, in turn, can be plugged into equation (6) to compute $P(z_i = 1)$.[2]

[2b] $z_i = 0$ means customer i dropped out after the last purchase at $t_{x(i)}$ before T_i. From the likelihood function (5) conditioned on $z_i = 0$, $\lambda = \lambda_i$ and $\mu = \mu_i$, therefore, y_i follows

[2] To facilitate convergence by allowing even drawing from the parameter space of μ, it sometimes helps to multiply the expression (6) by a prior that imposes an upper bound on the probability for z=1. This is because, when z=1, MLE of μ is 0, implying no dropout.

an exponential distribution with parameter $\lambda_i+\mu_i$ and truncation such that $t_{x(i)} < y_i < T_i$.[3]

[2c] Given z_i and y_i, equation (5) 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-Hastings algorithm is used to generate λ_i first and then μ_i, where the proposal distribution is chosen to be lognormal.

[3] See Bayesian textbooks elsewhere for details on the multivariate normal regression update (Congdon 2001; Gelman et al. 1995; Rossi, Allenby, and McCulloch 2005).

4. EMPIRICAL ANALYSIS

The appendix illustrates the simulation study to confirm the recovery of the predetermined parameters, λ_i and μ_i (i=1,..,400). We now apply the proposed model (hereafter denoted as the HB model) to three different types of databases --- music CD for e-commerce, FSP data for a department store and a music CD chain --- and make comparison with the Pareto/NBD model.

4.1. E-commerce data for CDNOW

The first dataset is CDNOW data used by Fader, Hardie and Lee (2005a, b), kindly made available for our study. The database contains e-commerce transactions over 78 weeks (1/1/97~6/30/98) at the CDNOW website. It includes data on 2357 customers who became a member of CDNOW during the first 12 weeks. As in Fader et al., the first 39

[3] The author appreciates Siddharth S. Singh, Sharad Borle, and Dipak C. Jain for pointing out the fact that the conditioning of λ was missing from this distribution in the earlier version of the manuscript.

weeks of the data are used for model calibration and the second 39 weeks are used for model validation. Thus, depending on when customers became a member, the length of the observation (T) in the calibration sample varies from 27 to 39 weeks. Note that almost 60% of the customers (i.e., 1411 customers) make no repeat purchases what so ever during the calibration period (i.e., x=0 and only an initial purchase), which makes the estimation challenging. Because the dataset did not contain any customer demographic data, we used a dollar amount of the initial purchase as the only covariate. Table 1 summarizes the descriptive statistics of the calibration sample.

< Insert Table 1 around here >

The MCMC steps were repeated 15,000 iterations, of which the last 5000 were used to infer the posterior distribution of the parameters. Convergence was monitored visually and checked with the Geweke test (Geweke 1992). 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 mean squared errors (MSE) between predicted and observed numbers of transactions for individual customers were used. For an aggregate measure, mean absolute percent errors (MAPE) between predicted and observed weekly cumulative transactions was used. Table 2 compares the result along with that of the intercept-only HB model (M1). Both Pareto/NBD and HB M2 models provide similar fit, however, the former seems to perform better, especially in aggregate tracking. This can be seen from the time-series tracking of the cumulative number of transactions in Figure 2. The vertical dotted line at Week 39 separates the validation from the calibration period.

< Insert Table 2 and Figure 2 around here >

For a visual check at the disaggregate level, Figure 3 shows the predicted number of transactions during the validation period averaged across individuals, conditional on the number of transactions made during the calibration period. This measure was used by Fader et al. (2005a, b) as well.

< Insert Figure 3 and Table 3 around here >

Table 3 reports the estimation result of the HB models for the posterior means of the parameters and their 2.5 and 97.5 percentiles in the parentheses as a standard-error like measure. The marginal loglikelihood suggests that M2 is better than M1. Note that the left hand side of the regression is a logarithm of λ and μ, and the magnitude of the coefficients must be interpreted accordingly. The only covariate, the amount of an initial purchase, is significantly positive for purchase frequency. This implies that customers with a larger initial purchase (in dollars) tend to shop more often. Lifetime does not seem to be related to the amount so that customers with large and small initial purchases are equally likely to remain active.

< Insert Figure 4 around here >

Figure 4 is a scatter plot of the posterior means of λ_i and μ_i (i=1,…,2357) for individual customers. The pattern does not suggest any particular relationship between the two parameters. This observation is also supported from the fact that, for the intercept-only model, the correlation between $\log(\lambda)$ and $\log(\mu)$ was 0.05, according to the estimated variance-covariance matrix (Γ_0) of the normal mixture distribution.[4] The

[4] To check whether the independence assumption of Pareto/NBD is satisfied, correlation of Γ_0 must be tested on the intercept-only model but not the covariate model. This is because, if covariates explain the correlation between λ and μ completely, then no

estimate is not significantly different from 0, which can be confirmed from Figure 5, the distribution of the estimated correlation for M1 obtained from a byproduct of the 5000 MCMC iterations. Both the scatter plot and the histogram imply that the independence assumption of λ and μ in the Pareto/NBD model holds for this dataset.

< Insert Figure 5 and Table 4 around here >

Table 4 reports six customer-specific statistics: the posterior means of λ_i and μ_i along with their 2.5% and 97.5% quantiles,[5] an expected lifetime after the last transaction, a survival rate after one year, the probability of being active at the end of the calibration period, and the expected number of transactions during the validation period, for the best and worst 10 customers with respect to the last statistic. These statistics are useful in actual CRM, for example, by providing customer ranking. The last three rows of Table 4 report the average, minimum, and maximum of the six statistics for the 2357 customers. For example, the probability of being active at the end of the calibration period (39th week) varies from 0.000 to 1.000 with the average being 0.425. The expected number of transactions during the validation period ranges from 0.00 to 22.59 with the average being 0.63.

Because the HB model produces a complete distribution of λ_i and μ_i at the customer level as a byproduct of the MCMC method $\{\lambda_i^{(g)}$ and $\mu_i^{(g)}$; $g=1,..,5000; i=1,..,2357\}$, it is a simple algebraic computation to obtain the distribution of any statistic that is a function

---

correlation remains in that Γ_0.

[5] The implied 95% confidence intervals for the individual-level parameters λ_i and μ_i are wide because the likelihood involves only the RF data from that customer. The Pareto/NBD model is expected to produce even wider 95% confidence intervals, whereby their rough magnitude can be inferred from the variance of the posterior gamma distributions shown in Section 3.1.

of λ and μ. Statistics in the last four columns are obtained using equations (9), (10), (6), and (8), respectively, in such a straightforward manner.[6] In contrast, the last two statistics are claimed to be the main result by SMC through a complex derivation (equations (11)~(13), (22) and the appendix in their paper), and their computation involves non-standard hypergeometric functions of various kinds. The fact that a distribution rather than a point estimate of a statistic is obtained also permits the application of statistical inference at the individual level.

Customer 2357 is interesting because she made 21 repeat purchases during the first 4.7 weeks and then stopped buying thereafter. This is the reason for her high purchase rate λ_i, the low probability of being active at the end of the calibration period, and the small expected number of transactions in the validation period. Managerially, the ability of the proposed model to provide information on individual customers with ease, is especially suited to CRM application.

4.2. FSP data for a department store

This dataset contains shopping records for the members of a frequent shopper program (FSP) at a department store in Japan. It sells a wide variety of merchandise ranging from apparel, interior decoration, electronics, toys, to gourmet food on over ten floors. The period of data covers for 52 weeks starting from July 1, 2000.

We drew a random sample of 400 customers who had joined the FSP during the month of July 2000, and hence, their shopping data were available from their first purchase as a member until June 29, 2001. The first and the last 26 weeks of the data were used for calibration and validation, respectively. Multiple receipts at different

[6] Some of these equations appear in the online appendix.

cashiers within the same day were consolidated as one spending, and returns (negative spending, credit, etc.) were ignored. The descriptive statistics for the calibration period are shown in Table 5. The number of repeat purchases x_i ranges from 0 for 17 customers to 101, implying that some customers made purchases almost every other day.[7] Inspection of the distributions of interpurchase times within customers reveals that our dataset appears to satisfy the Poisson assumption.

< Insert Table 5 about here >

The database contains very limited information on customer characteristics: gender, age, and address. Because majority of the shoppers visit the store by means of public transportation and through transit during commuting, a geographical distance between home address and the store does not necessary correspond to the store accessibility. Instead, we constructed a covariate called FOOD, the fraction of store visits on which food items were purchased, as the proxy for accessibility.[8] Thus, FOOD ranges from 0 to 1: 1 if foods were purchased on every shopping occasion, 0.5 if purchased on a half of the shopping occasions, and 0 if no foods were purchased on any of the shopping occasions. Another covariate, average spending per trip, was created from the customer purchase records. To keep the scales comparable among the covariates, the unit for the average spending was defined as the 10^{-5} yen and AGE as the one-hundredth of the actual age.

< Insert Table 6 about here >

Table 6 compares the aggregate and disaggregate fits of the two models in the calibration and validation samples. Both models provide similar fit, however, the HB

[7] This is hardly surprising because the department store is directly connected to a busy train hub station and easily accessible through transit during commute.

[8] We use the terms "purchase" and "visit" interchangeably because store visit by a customer can be verified only through his/her purchase.

seems to perform slightly better for this dataset. This fact can be seen from a weekly tracking of the cumulative number of transactions in Figure 6. Figure 7, the conditional expectation of future transactions given the number of transactions made during the calibration period, also results in similar fits.

< Insert Figures 6 and 7 and Table 7 around here >

Table 7 reports the estimation result of the full HB model (M3) along with nested models that include only an intercept (M1) and two covariates (M2). For the intercept-only model, correlation between the purchase and dropout rates is not significantly different from 0, satisfying the assumption of the Pareto/NBD model. This is also evident from the scatter plot of the posterior means of the individual level parameters, λ_i and μ_i, shown in Figure 8.

< Insert Figure 8 around here >

As covariates are added, the estimated coefficients remain stable and the marginal loglikelihood increases, where the full HB is chosen as the best model. Significant covariates for the purchase rate are average spending and FOOD, the fraction of store visits on which food items were purchased and a proxy for store accessibility. Age and gender do not affect the purchase rate. For the dropout rate, none of the covariates is significant. Customer lifetime does not differ by the amount of average spending, food/non-food buyers, age, or gender.

This implies that food buyers and small spenders visit the store more often: the finding consistent with the story told by the store manager. Although food buyers spend a smaller amount on each shopping trip, they visit the store often enough such that the retailer considers food shoppers as their key clients. Due to lower margins on food items relative to other categories, such as jewelry, fashion items, and interior decoration, the food section in itself is not particularly profit making. Yet, the retailer recently renovated the entire food floor featuring fancy gourmet and imported foods, in an effort to attract this target segment.

4.3. Retail FSP data for a Music CD chain

The third dataset was obtained from a FSP of a large chain for music CD. The period covered for 52 weeks starting from September 1, 2003. We extracted 500 random customers who had joined the FSP with the initial purchase (trial) during the first quarter and also at least one repeat purchase during the first half. The first 26 weeks and the last 26 weeks of the data were used as calibration and validation samples, respectively. The available customer characteristics were the amount of the initial purchase, age, and gender. The descriptive statistics of the calibration sample are shown in Table 8.

< Insert Tables 8 and 9 and Figure 10 about here >

Table 9 compares the aggregate and disaggregate fits of the two models in the calibration and validation samples. Figure 9 is a weekly tracking of the cumulative number of transactions, and Figure 10 shows the conditional expectation of future transactions given the number of transactions made during the calibration period. The HB model seems to perform slightly better than the Pareto/NBD model.

< Insert Table 10 and Figure 11 around here >

Table 10 shows the estimated coefficients of the nested HB models, in which the full model (M3) has the best marginal loglikelihood. For the intercept-only model (M1), correlation between the purchase and dropout rates is not significantly different from 0, just like the previous two datasets. This can be seen from a scatter plot of the posterior means of λ and μ for the 500 customers in Figure 11. The only significant parameter is the amount of initial purchase on the purchase rate. It implies that customers with a larger spending on their first purchase tend to visit the chain more often. No difference in terms of age and gender is found on purchase frequency. The amount of initial purchase, age, and gender are not related to customer lifetime.

5. CONCLUSIONS

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 change, the Pareto/NBD model of customer-base analysis was updated, resulting in an hierarchical Bayes model, which was then estimated by an MCMC method.

The HB model presumes three tried and tested assumptions of Pareto/NBD: (1) a Poisson purchase process, (2) a memoryless dropout process (i.e., constant hazard rate), and (3) heterogeneity across customers, while relaxing Pareto/NBD's independence assumption of the purchase and dropout processes. Because customer heterogeneity is captured as a prior in a hierarchical Bayesian framework rather than through a mixture distribution, the entire modeling effort can bypass all the complications associated with aggregation, which is left to MCMC simulation.

The HB model was shown to perform well in the empirical analysis using three datasets of various types. Outputs, obtained from by-products of the MCMC estimation, 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. These customer-specific statistics can be quite useful, for example, for ranking customers, in actual CRM.

The simplicity of the HB model has lead to an estimable model, in which λ and μ are a function of customer characteristic variables. Such models demonstrate that recency-frequency data, in conjunction with customer behavior and characteristics, can provide important insights into direct marketing issues, such as the demographic profile of best customers and whether long-life customers spend more.

The current study also confirmed that, in all three datasets, the independence assumption of the purchase and dropout rates for the Pareto/NBD model, which, in turn, provided sound performance in fit and prediction. 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 statistical inference on the correlation between $\log(\lambda)$ and $\log(\mu)$, obtained from Γ_0 of an intercept-only model.

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 customized 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. Though we did not find an evidence for the relationship between lifetime and average spending from our department store data, the current study is more methodological in nature and falls short of drawing any substantive conclusions. Pioneering research by Reinartz and Kumar (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 Kumar (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 seems 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 transactions to dollar amounts 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 Kumar (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 Len 1999). A Poisson process implies random purchase occurrence (constant hazard rate) 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 dynamic purchase pattern could provide valuable insights into CRM. One approach is to incorporate time-varying covariates as a function of log(μ), which results in a time-varying hazard function (Gonul and Hofstede 2006). This extension, however, puts more burdens on the part of data collection, because the model estimation requires not just recency and frequency but the complete purchase history.

ACKNOWLEGDMENTS

Makoto Abe appreciates the comments and suggestions of the reviewers, area editor and editor. He also benefited greatly from discussions with Don Morrison and Randy Bucklin during his sabbatical stay at UCLA. The author thanks Pete Fader and Bruce Hardie for useful comments and making the CDNOW data available. In addition, he acknowledges the 2004 Data Analysis Competition sponsored by the Operation Research Society of Japan for the use of the department store data, and the financial support from the Japanese Ministry of Education Scientific Research Grant.

APPENDIX: SIMULATION STUDY

A simulation study was conducted to examine the recovery of parameters. RF data were generated according to the Poisson purchase and exponential dropout assumptions from pre-specified parameters. (λ_i, μ_i) (i=1,..,400) were generated from a lognormal distribution of (4).

$$\begin{bmatrix} \log(\lambda) \\ \log(\mu) \end{bmatrix} \sim MVN\left(\theta_0 = \begin{bmatrix} -3 \\ -6 \end{bmatrix}, \Gamma_0 = \begin{bmatrix} 0.5 & -0.16 \\ -0.16 & 1.0 \end{bmatrix} \right)$$

The parameters for the lognormal reflect the estimated values from actual data. This corresponds to correlation between log(λ) and log(μ) of -0.2263 (=-0.16/sqrt(0.5×1.0)).

To account for the sample variation in data generation, the simulation was repeated 30 times. Each simulation was put through 15,000 MCMC steps, of which the last 5,000 draws were used to construct the posterior estimation.

Figures 1A shows the scatter plot of the true and estimated values for the 400 customers from a simulation sample. Correlation between the true and posterior mean of the 400 customers averaged over 30 simulation samples are 0.80 and 0.18 for λ and μ, respectively. Though the recovery of λ is fine, the marginal result for μ arises from the following reasons: (1) customer dropout is never observed directly and we are trying to estimate the latent parameter, (2) the sample size is one RF data point per customer, and (3) point estimate (posterior mean) is compared against the true value.

< Insert Figure 1A around here >

Because correlation is sensitive to outliers, perhaps a better measure that fully takes advantage of the HB method with an exact small-sample standard error is to count how many true values lie outside the 95% confidence interval. Theoretically, it should be approximately 5% of the cases or 20 in our case. The average of 30 simulation samples results in 18.5 and 15.8 for λ and μ, respectively, which is reasonable. In summary, this simulation study shows that the true underlying parameters can be recovered with reasonable accuracy.

The difficulty in estimating individual lifetime parameters μ_i arises from the fact that the individual likelihood involves just two data points, RF data for that customer. Hence, inability to accurately estimate μ_i is due to this lack of degrees of freedom rather than model limitation. This can be verified by computing a confidence interval of the individual-level parameter μ_i for a standard Pareto/NBD model, as stated in Footnote 5.

As shown in Section 3.1., when the prior distribution of μ_i is *gamma*(s, β), the posterior is either *gamma*(s+1, β+y_i) or *gamma*(s, β+T_i) depending on whether the customer has died at y_i<T_i or is alive at T_i. By positing the largest possible value for T_i and y_i, one can obtain the lower bound for the standard error of the gamma. The table below shows the result for CDNOW data, in which s ≅ 0.6 and β ≅ 12 (FHL 2005).

gamma mixture for μ_i		**mean**	**std. error (σ)**	**4×σ ≅ 95%CI**
prior		0.05	0.065	0.26
posterior	**dead by T_i**	0.036	0.028	0.112
	alive at T_i	0.013	0.017	0.068

It implies that the 95% confidence interval of μ_i for the Pareto/NBD should be larger than 0.068~0.112, which is comparable to the figures reported in Table 4.

ONLINE APPENDIX

Derivation of Survival Probability and Likelihood Function

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

$$P(\tau > T \mid \lambda, \mu, x, t, T) = P(alive \mid history)$$

$$= \frac{P(alive \,\&\, history)}{P(history)} \tag{7}$$

$$= \frac{P(history \mid alive)\, P(alive)}{P(alive \,\&\, history) + P(dead \,\&\, history)}$$

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

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

Furthermore, the following two equations can be derived.

$$\begin{aligned} P(history \mid alive) &= P(x - \text{th purchase at } t \,\&\, \text{nopurchase between } [t_x, T]) \\ &= \frac{\lambda^x t_x^{\;x-1}}{\Gamma(x)} e^{-\lambda\, t_x} \times e^{-\lambda (T - t_x)} \\ &= \frac{\lambda^x t_x^{\;x-1}}{\Gamma(x)} e^{-\lambda T} \end{aligned}$$

$$\begin{aligned} P(history \,\&\, dead) &= \int_t^T P(x - \text{th purchase at } t_x \,\&\, \text{nopurchase between } [t_x, y] \,\&\, \text{die at } y \in [t_x, T])dy \\ &= \int_{t_x}^T \frac{\lambda^x t_x^{\;x-1}}{\Gamma(x)} e^{-\lambda\, t_x} \times e^{-\lambda (y - t_x)} \times \mu e^{-\mu y} dy \\ &= \frac{\lambda^x t_x^{\;x-1}}{\Gamma(x)} \mu \int_{t_x}^T e^{-(\lambda+\mu)\,\mathrm{y}} dy \\ &= \frac{\lambda^x t_x^{\;x-1}}{\Gamma(x)} \frac{\mu}{\lambda + \mu} \left\{ e^{-(\lambda+\mu)\, t_x} - e^{-(\lambda+\mu)\,\mathrm{T}} \right\} \end{aligned}$$

Substituting the three equations above into equation (7) leads to the survival probability formula.

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

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

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

$$E[X(t) \mid \lambda, \mu] = \lambda E[\eta] = \frac{\lambda}{\mu}\left(1 - e^{-\mu t}\right) \qquad \text{where } \eta = \min(\tau, t) . \tag{8}$$

Formulas for other relevant individual statistics are

$$\text{The expected lifetime} = \frac{1}{\mu} \tag{9}$$

$$\text{The survival rate after 1 year} = \exp(-52\mu) \qquad \text{where time unit is expressed in weeks} \tag{10}$$

Derivation of the Bayesian Update for a Pareto/NBD Model in Section 3.1

If active at T_i, the likelihood is $\frac{\lambda^x t_x^{\ x-1}}{\Gamma(x)} e^{-(\lambda+\mu)T}$ as shown in section 3.1 for case z=1.

Recall the priors, $g(\lambda \mid \alpha, r) = \frac{\alpha^r \lambda^{r-1}}{\Gamma(r)} e^{-\alpha\lambda}$, $\quad g(\mu \mid \beta, s) = \frac{\beta^s \mu^{s-1}}{\Gamma(s)} e^{-\beta\mu}$.

For the posterior for λ, μ can be considered as a constant. Thus,

$$g(\lambda \mid \alpha, r, x, t, T) \propto \lambda^x e^{-(\lambda+\mu)T} \lambda^{r-1} e^{-\alpha\lambda} \propto \lambda^{x+r-1} e^{-(\alpha+T)\lambda} \sim gamma(r+x, \alpha+T) .$$

Likewise, λ can be considered as a constant for the posterior of μ, and hence,

$$g(\mu \mid \beta, s, x, t, T) \propto e^{-(\lambda+\mu)T} \mu^{s-1} e^{-\beta\mu} \propto \mu^{s-1} e^{-(\beta+T)\mu} \sim gamma(s, \beta+T) .$$

If inactive at T_i, a similar derivation results in the desired update formula.

REFERENCES

Allenby, Greg M., 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.

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

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.

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.

Gönül, Füsun F. and Frenkel Ter Hofstede (2006), "How to Compute Optimal Catalog Mailing Decisions", *Marketing Science*, 25 (1), 65–74.

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.

Rust, Roland T. and Tuck Siong Chung (2006), "Marketing Models of Service and Relationships", *Marketing Science*, 25 (6), 560–580.

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.

Sun, Baohong (2006), “Technology Innovation and Implications for Customer Relationship Management”, *Marketing Science*, 25 (6) 594–597.

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.

Table 1. Descriptive Statistics for CDNOW Data

	mean	std. deviation	min	max
Number of repeats	1.04	2.19	0	29
Observation duration T (days)	229.01	23.29	189	272
Recency, (T-t) (days)	181.09	77.11	0	272
Amount of initial purcahse ($)	32.99	34.66	0	506.97

Table 2. Model Fit for CDNOW Data

Criterion		**Pareto/NBD**	**HB M1** (no covariates)	**HB M2** (with a covariate)
Disaggregate Measure				
Correlation	**validation**	0.63	0.62	0. 62
	calibration	1.00	0.98	0.98
MSE	**validation**	2.57	2.61	2.62
	calibration	0.64	0.58	0.58
Aggregate Measure				
Timeseries MAPE (%)	**validation**	1.3	2.6	2.4
	calibration	8.9	7.5	7.6
	pooled	5. 1	5.6	5.4

Table 3. Estimation Result for CDNOW Data

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

		HB M1 (no covariates)	**HB M2** (with a covariate)
Purchase rate log(λ)	**Intercept**	-3.53 (-3.76, -3.35)	-3.74 (-3.91, -3.56)
	Initial amount ($ 10^{-3})	---	3.21* (1.59, 4.90)
Dropout Rate log(μ)	**Intercept**	-3.56 (-4.05, -3.27)	-3.62 (-4.03, -3.34)
	Initial amount ($ 10^{-3})	---	-0.21 (-2.52, 1.98)
$\sigma_\lambda^2 \equiv \mathrm{var}[\log \lambda]$		1.33 (1.07,1.72)	1.41 (1.14, 1.70)
$\sigma_\mu^2 \equiv \mathrm{var}[\log \mu]$		2.56 (1.60, 4.66)	1.59 (0.83, 3.47)
$\sigma_{\lambda\mu} \equiv \mathrm{cov}[\log \lambda, \log \mu]$		0.11 (-0.26, 0.68)	0.06 (-0.24, 0.57)
correlation computed from Γ_0		0.05 (-0.16, 0.30)	0.03 (-0.18, 0.26)
marginal loglikelihood		-1381	-1360

Table 4. Customer-Specific Statistics for the Top and Bottom 10 Customers

ID	mean(λ)	2.5% tile	97.5% tile	mean(μ)	2.5% tile	97.5% tile	Mean Expected lifetime (years)	1 year survival rate	Probability of being active at the end of calibration	Expected number of transactions in validation period
1	0.778	0.531	1.069	0.0187	0.0021	0.0523	1.46	0.487	0.997	22.59
2	0.681	0.498	0.914	0.0162	0.0022	0.0432	1.62	0.522	0.995	20.33
3	0.510	0.318	0.785	0.0183	0.0024	0.0510	1.41	0.493	0.991	14.83
4	0.497	0.304	0.684	0.0172	0.0023	0.0478	1.60	0.509	0.997	14.72
5	0.444	0.294	0.612	0.0174	0.0024	0.0495	1.53	0.504	0.986	12.92
6	0.397	0.227	0.595	0.0204	0.0023	0.0563	1.28	0.467	0.967	11.02
7	0.381	0.216	0.601	0.0181	0.0026	0.0522	1.46	0.491	0.978	10.92
8	0.330	0.202	0.480	0.0175	0.0021	0.0481	1.42	0.497	0.992	9.64
9	0.313	0.188	0.493	0.0174	0.0021	0.0492	1.56	0.504	0.989	9.14
10	0.301	0.167	0.435	0.0170	0.0022	0.0469	1.49	0.505	0.998	8.91
...	**...**	**...**	**...**	**...**	**...**	**...**	**...**	**...**	**...**	**...**
2348	0.389	0.054	1.093	0.1150	0.0108	0.3959	0.31	0.140	0.045	0.10
2349	0.524	0.244	0.888	0.0573	0.0084	0.1560	0.48	0.199	0.014	0.09
2350	0.277	0.051	0.692	0.0889	0.0079	0.3002	0.37	0.162	0.036	0.08
2351	0.394	0.061	0.943	0.1188	0.0109	0.4342	0.30	0.135	0.030	0.07
2352	0.334	0.104	0.686	0.0678	0.0111	0.1878	0.40	0.162	0.014	0.05
2353	0.372	0.094	0.875	0.0779	0.0095	0.2490	0.40	0.165	0.017	0.04
2354	0.448	0.138	0.888	0.0791	0.0106	0.2314	0.38	0.149	0.003	0.01
2355	0.844	0.138	2.358	0.1205	0.0118	0.4231	0.28	0.118	0.004	0.01
2356	0.712	0.304	1.321	0.0854	0.0108	0.2564	0.33	0.139	0.001	0.00
2357	3.535	2.435	4.771	0.0936	0.0107	0.2663	0.31	0.138	0.000	0.00
ave	0.054	0.013	0.138	0.0620	0.0037	0.2066	0.75	0.316	0.425	0.63
min	0.024	0.002	0.067	0.0158	0.0015	0.0422	0.28	0.118	0.000	0.00
max	3.535	2.435	4.771	0.1314	0.0118	0.4929	1.74	0.536	1.000	22.59

Table 5. Descriptive Statistics for Department Store FSP Data

	mean	std. deviation	min	max
Number of repeats	16.02	16.79	0	101
Observation duration T (days)	171.24	8.81	151	181
Recency, (T-t) (days)	24.94	42.82	0	181
Average spending ($\times 10^5$ yen)	0.067	0.120	0.0022	1.830
FOOD	0.79	0.273	0.0	1.0
AGE	52.7	14.6	22	87
FEMALE	0.93	0.25	0	1

Table 6. Model Fit for Department Store FSP Data

Criterion		Pareto/NBD	HB M3 (with 4 covariates)
Disaggregate Measure			
Correlation	**validation**	0.90	0. 90
	calibration	1.00	1.00
MSE	**validation**	58.2	57.8
	calibration	1.22	3.6
Aggregate Measure			
Timeseries MAPE (%)	**validation**	2.29	5.09
	calibration	18.2	16.2
	pooled	10.3	10.6

Table 7. Estimation Result for Department Store FSP Data

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

		HB M1 (no covariates)	**HB M2** (with 2 covariates)	**HB M3** (with 4 covariates)
Purchase rate **log(λ)**	**Intercept**	-0.83 (-93, -0.72)	-1.64 (-2.00, -1.28)	-1.72 (-2.23, -1.21)
	Average spending	---	-0.22* (-0.36, -0.09)	-0.23* (-0.37, -0.10)
	FOOD	---	1.22* (0.83, 1.60)	1.22* (0.81, 1.63)
	AGE	---	---	-0.03 (-0.66, 0.58)
	FEMALE	---	---	0.11 (-0.23, 0.45)
Dropout Rate **log(μ)**	**Intercept**	-6.26 (-7.02, -5.66)	-4.83 (-6.25, -3.62)	-4.75 (-6.35, -3.29)
	Average spending	---	-0.30 (-1.17, 0.24)	-0.54 (-1.37, 0.12)
	FOOD	---	-1.48 (-2.86, 0.01)	-1.57 (-3.06, 0.05)
	AGE	---	---	-0.11 (-2.06, 1.74)
	FEMALE	---	---	0.20 (-0.96, 1.50)
$\sigma_\lambda^2 \equiv \mathrm{var}[\log \lambda]$		0.88 (0.73, 1.05)	0.70 (0.58, 0.84)	0.69 (0.57, 0.83)
$\sigma_\mu^2 \equiv \mathrm{var}[\log \mu]$		1.96 (0.82, 4.08)	1.62 (0.66, 3.22)	1.49 (0.60, 2.83)
$\sigma_{\lambda\mu} \equiv \mathrm{cov}[\log \lambda, \log \mu]$		-0.43 (-0.91, 0.03)	-0.30 (-0.66, 0.04)	-0.30 (-0.64, 0.01)
correlation computed from Γ_0		-0.32 (-0.60, 0.03)	-0.29 (-0.54, 0.04)	-0.29 (-0.54, 0.01)
marginal loglikelihood		-1917	-1913	-1904

Table 8. Descriptive Statistics for Music CD Chain FSP Data

	mean	std. deviation	min	max
Number of repeats	2.65	2.36	1	22
Observation duration T (days)	146.66	25.84	92	182
Recency, (T-t) (days)	52.65	40.99	1	172
Average spending ($\times 10^4$ yen)	0.359	0.198	0.095	2.048
AGE	31.5	9.8	7	78
FEMALE	0.49	0.50	0	1

Table 9. Model Fit for Music CD Chain FSP Data

Criterion		**Pareto/NBD**	**HB M3 (with 3 covariates)**
Disaggregate Measure			
Correlation	**validation**	0.59	0. 61
	calibration	0.95	0.95
MSE	**validation**	6.43	5.11
	calibration	2.14	1.70
Aggregate Measure			
Timeseries MAPE (%)	**validation**	11.81	4.13
	calibration	9.38	13.50
	pooled	10.60	8.82

Table 10. Estimation Result for Music CD Chain FSP Data

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

		HB M1 (no covariates)	**HB M2** (with 1 covariates)	**HB M3** (with 3 covariates)
Purchase rate $\log(\lambda)$	**Intercept**	-2.14 (-2.22, -2.06)	-2.27 (-2.40, -2.15)	-2.17 (-2.40, -1.94)
	Initial Purchase	---	0.38* (0.13, 0.63)	0.38* (0.13, 0.63)
	AGE	---	---	-0.14 (-0.75, 0.46)
	FEMALE	---	---	-0.12 (-0.27, 0.02)
Dropout Rate $\log(\mu)$	**Intercept**	-5.40 (-5.86, -5.04)	-5.50 (-6.11, -4.98)	-5.69 (-6.56, -4.85)
	Initial Purchase	---	0.20 (-0.83, 1.10)	0.21 (-0.90, 1.17)
	AGE	---	---	0.34 (-1.55, 2.19)
	FEMALE	---	---	0.17 (-0.45, 0.75)
$\sigma_\lambda^2 \equiv \mathrm{var}[\log \lambda]$		0.26 (0.20, 0.33)	0.25 (0.19, 0.31)	0.24 (0.18, 0.31)
$\sigma_\mu^2 \equiv \mathrm{var}[\log \mu]$		1.41 (0.79, 2.35)	1.46 (0.67, 2.47)	1.33 (0.67, 2.24)
$\sigma_{\lambda\mu} \equiv \mathrm{cov}[\log \lambda, \log \mu]$		0.11 (-0.04, 0.27)	0.10 (-0.05, 0.25)	0.10 (-0.04, 0.25)
correlation computed from Γ_0		0.18 (-0.08, 0.42)	0.16 (-0.08, 0.41)	0.18 (-0.08, 0.42)
marginal loglikelihood		-4425	-4424	-4417

Figure 1. Notations for RF Data

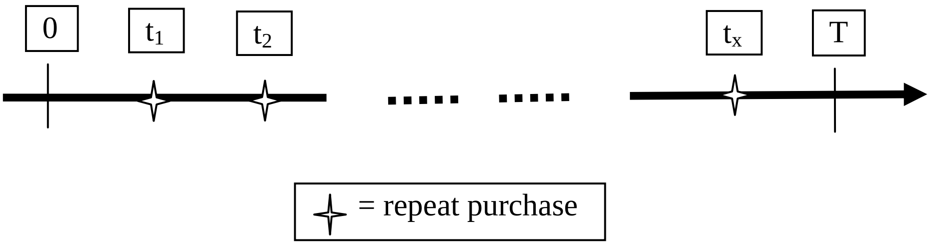

Figure 2. Weekly Time-series Tracking Plot for CDNOW Data

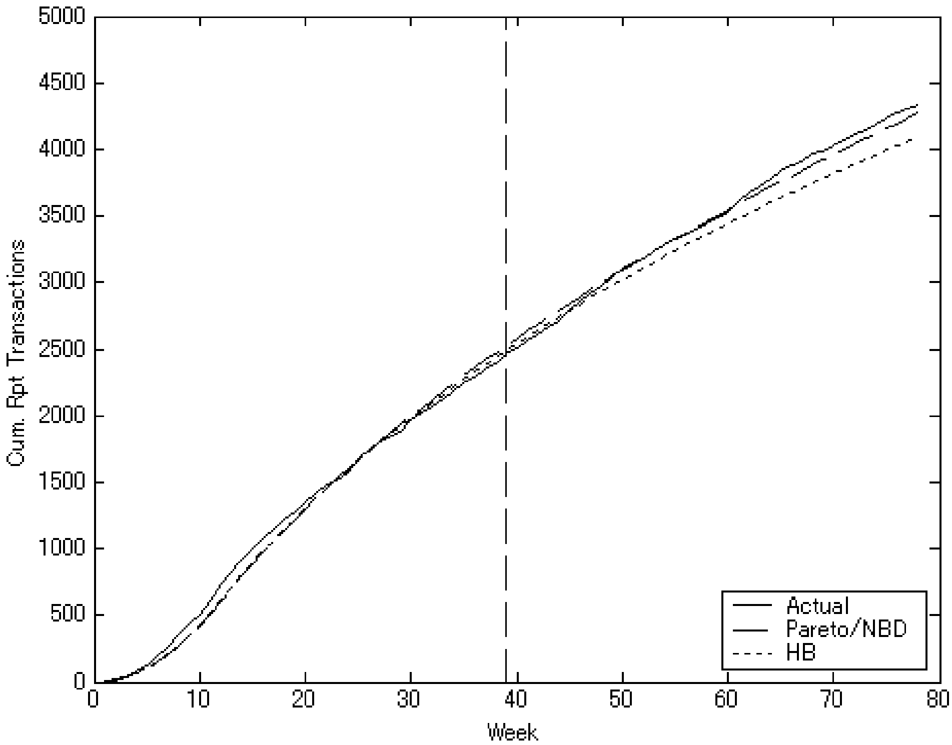

Figure 3. Conditional Expectation of Future Transactions for CDNOW

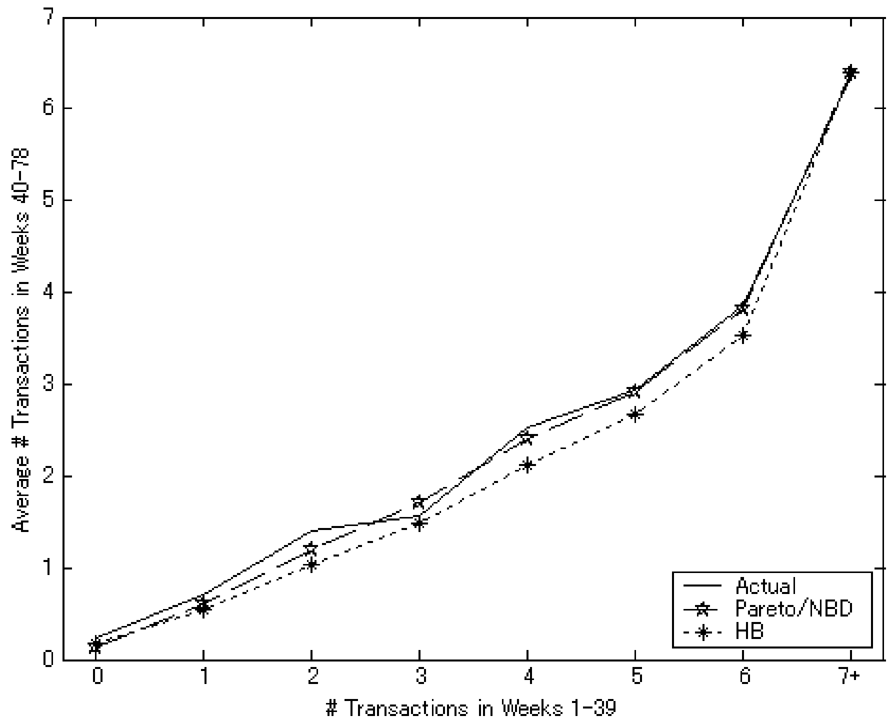

Figure 4. Scatter Plot of Posterior Means of λ and μ for CDNOW Data

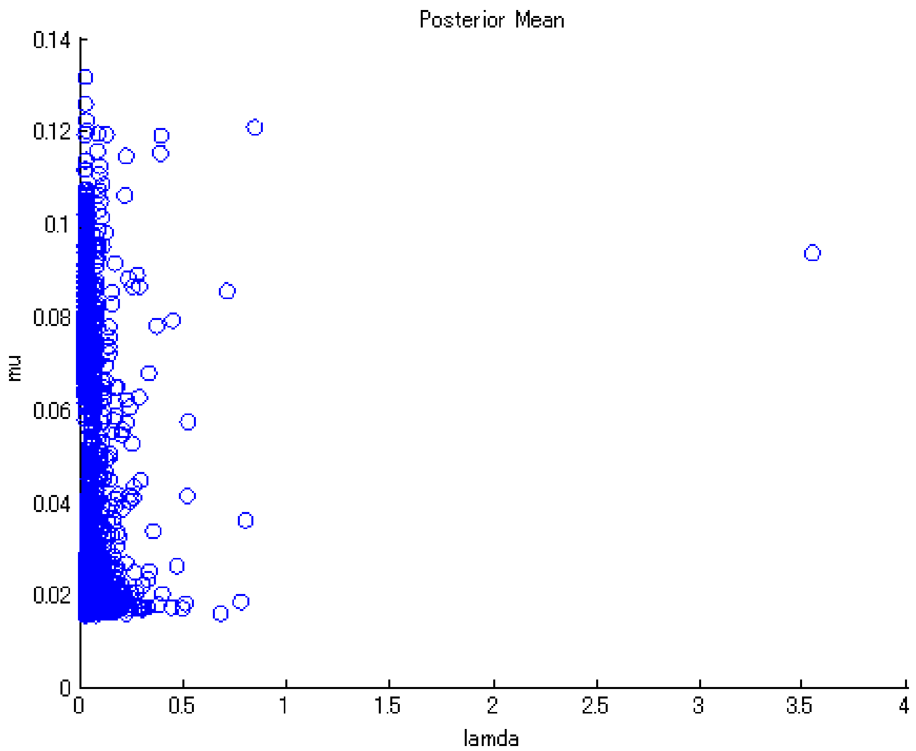

Figure 5. Distribution of Correlation between log(λ) and log(μ) for CDNOW Data

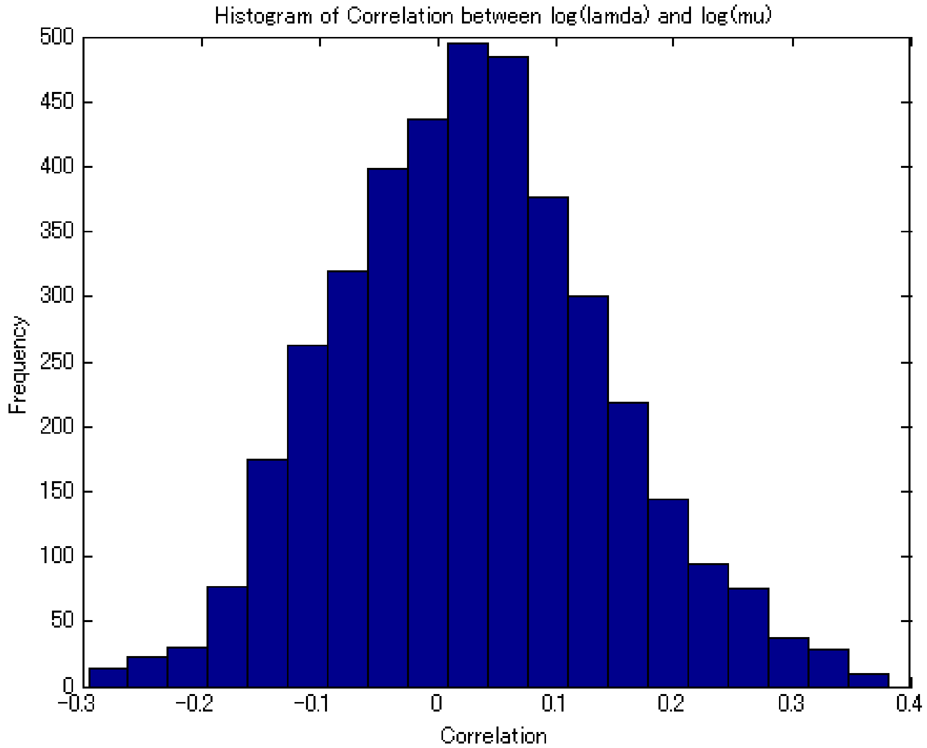

Figure 6. Weekly Time-series Tracking Plot for Department Store Data

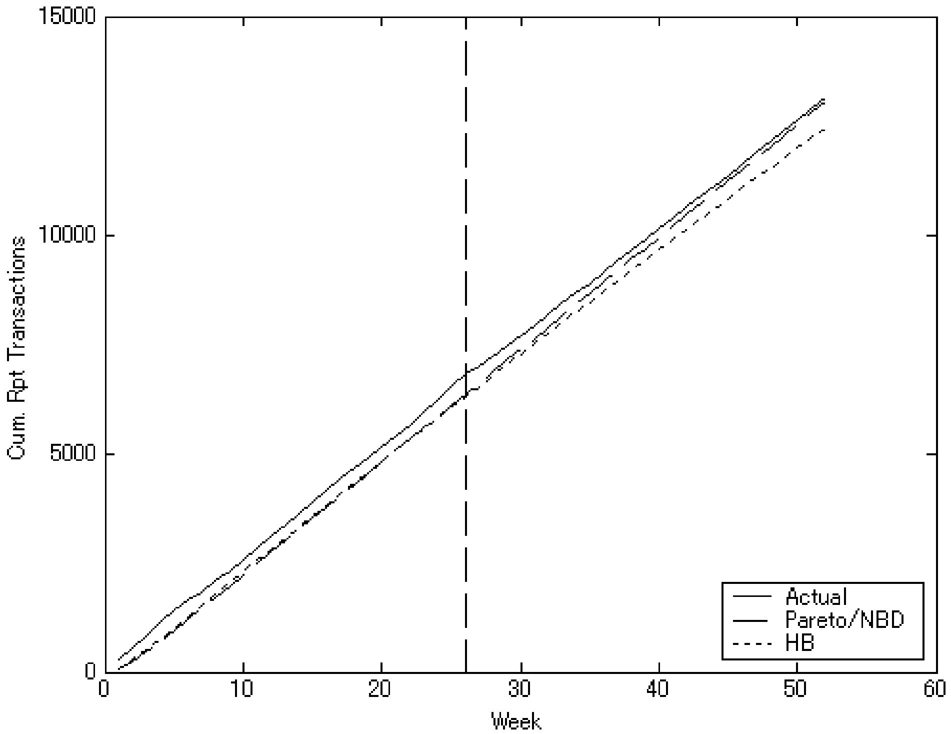

Figure 7. Conditional Expectation of Future Transactions for Department Store Data

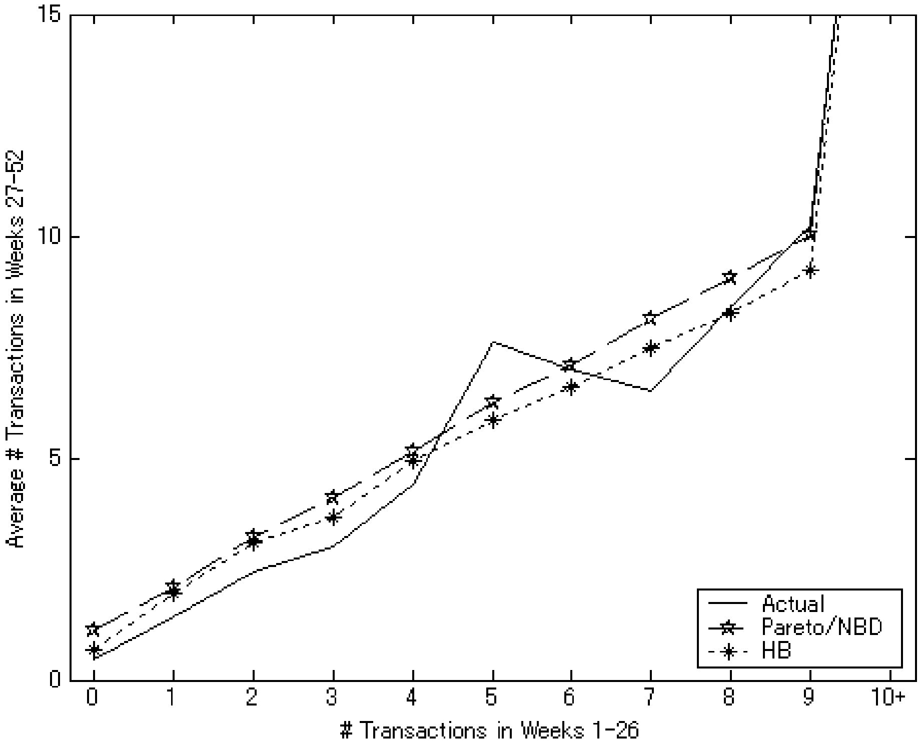

Figure 8. Scatter Plot of Posterior Means of λ and μ for Department Store Data

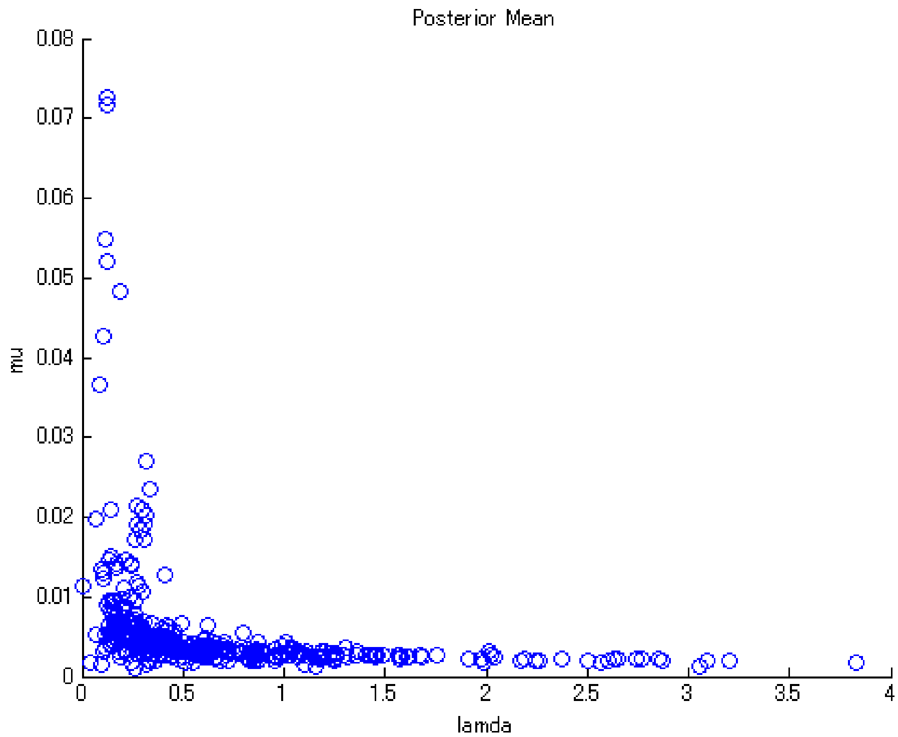

Figure 9. Weekly Time-series Tracking Plot for Music CD Chain Data

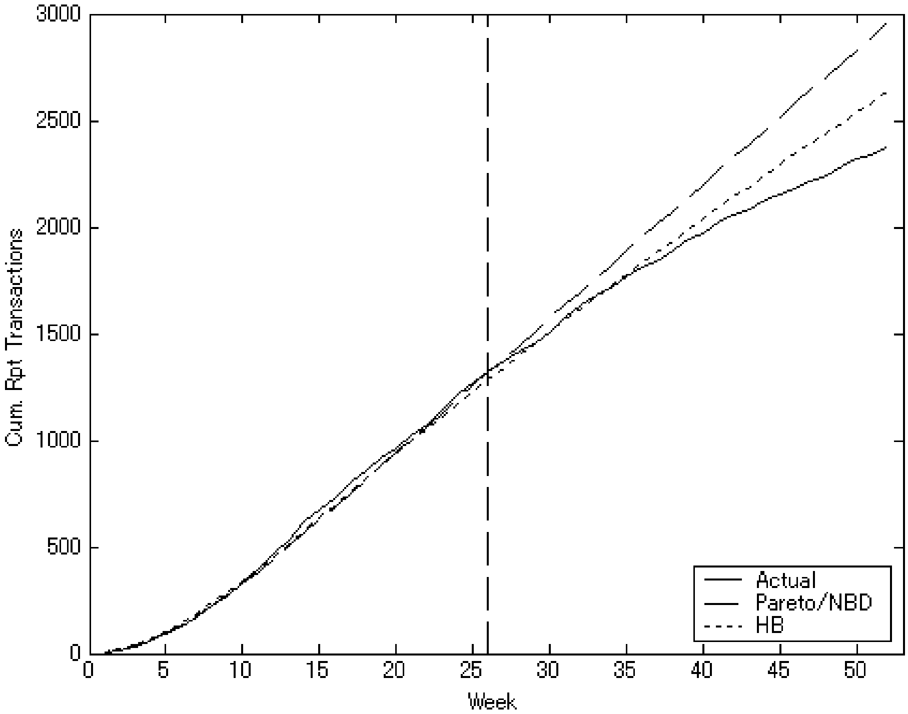

Figure 10. Conditional Expectation of Future Transactions for Music CD Chain Data

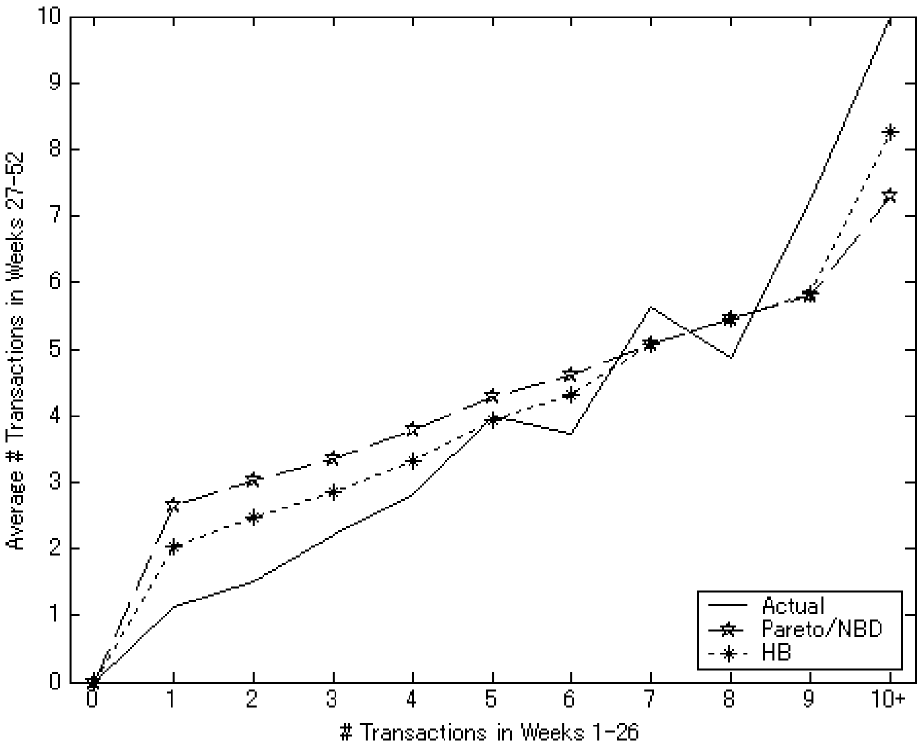

Figure 11. Scatter Plot of Posterior Means of λ and μ for Music CD Chain Data

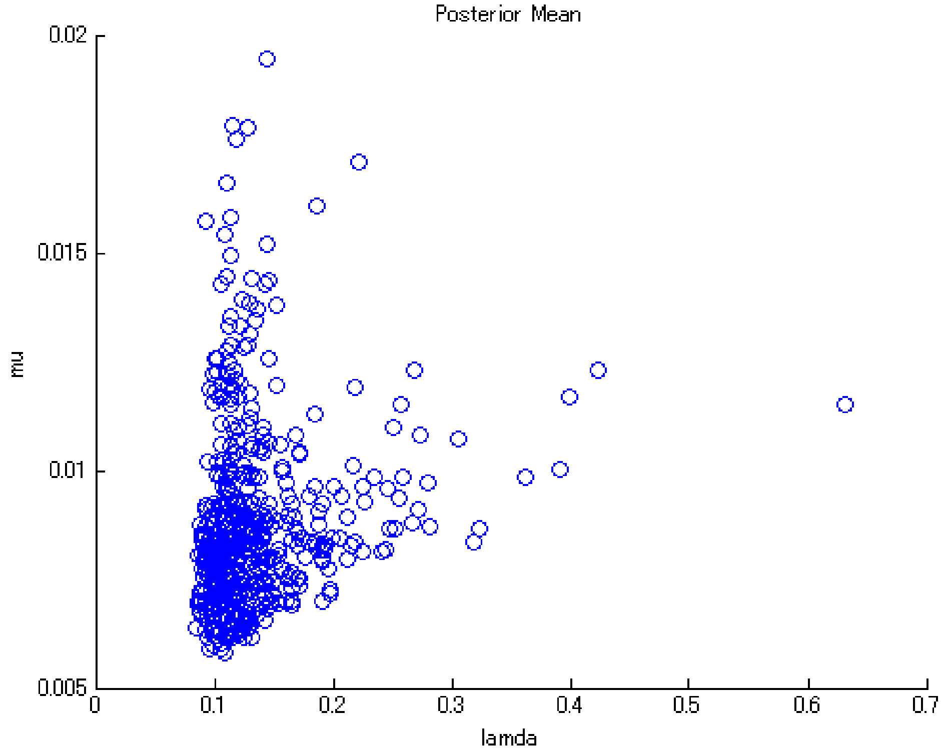

Figure A.1. Simulation Recovery of λ