INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
MEASURING THE VALUE OF CUSTOMERS: A HIERARCHICAL BAYES MODEL OF INTERPURCHASE TIME

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Lichung Jen, M.B.A. M.A.S.

* * * * *

The Ohio State University
1995

Dissertation Committee:  
Greg M. Allenby  
Robert P. Leone  
James L. Ginter

Approved by  
Advisor  
Business Administration
To My Wife and Sons
ACKNOWLEDGMENTS

I am most grateful to Professor Greg M. Allenby for his constant guidance and enduring encouragement. Thanks go to the other members of my advisory committee, Professors Robert P. Leone and James L. Ginter, for their valuable suggestions and comments.

I wish to thank Professor David C. Schmittlein for providing the data in this research. Part of financial assistance for this study was supported by William R. Davidson Doctoral Fellowship.

Finally, I am indebted to my wife and our parents and sisters, I deeply thank you for your understanding and unconditional support.
VITA

May 21, 1960 ........................................................... Born - Taipei, Taiwan, R.O.C.

June, 1983 ........................................................... B.A. in Business,
Chinese Culture University,
Taipei, Taiwan

June, 1985 ........................................................... M.B.A., Tamkang University,
Taipei, Taiwan

August 1991 ........................................................... M.S. in Applied Statistics
The Ohio State University,
Columbus, Ohio

June, 1987 - Sept. 1988 ......................................... Assistant Manager,
Wei-Chuan Foods Co.,
Taipei, Taiwan

Sept. 1988 - July 1989 ........................................... Assistant Research Fellow,
Chung-Hua Institution for
Economic Research,
Taipei, Taiwan

Oct. 1990 - Dec. 1991 ............................................. Statistical Consultant,
Nationwide Insurance Company,
Columbus, Ohio

Sept. 1991 - Present .............................................. Teaching and Research Associate.
Marketing Department, The Ohio
State University, Columbus, Ohio

FIELDS OF STUDY

Major Field: Business Administration

Minor Field: Marketing
# TABLE OF CONTENTS

| DEDICATION | ii |
| ACKNOWLEDGMENTS | iii |
| VITA | iv |
| LIST OF TABLES | vii |
| LIST OF FIGURES | viii |

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II. REVIEW OF THE LITERATURE</td>
<td>5</td>
</tr>
<tr>
<td>2.1 NBD Model</td>
<td>7</td>
</tr>
<tr>
<td>2.2 Hazard Function Models</td>
<td>9</td>
</tr>
<tr>
<td>2.3 Pareto-NBD model</td>
<td>11</td>
</tr>
<tr>
<td>2.4 The Importance of Individual Estimates</td>
<td>13</td>
</tr>
<tr>
<td>III. HIERARCHICAL BAYES MODEL</td>
<td>15</td>
</tr>
<tr>
<td>3.1 HB Model for Interpurchase</td>
<td>16</td>
</tr>
<tr>
<td>3.2 Robustness of the Restricted Inverse Gamma Distribution</td>
<td>24</td>
</tr>
<tr>
<td>3.3 Incorporating Covariates into the HB Model</td>
<td>27</td>
</tr>
<tr>
<td>3.4 HB Model for Monetary Value</td>
<td>32</td>
</tr>
<tr>
<td>IV. AN EMPIRICAL STUDY OF HIERARCHICAL BAYES MODEL FOR DIRECT MARKETING DATABASE</td>
<td>37</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Goodness of Fit of the Restricted Inverse Gamma (IG) Distribution</td>
<td>94</td>
</tr>
<tr>
<td>2. Goodness of Fit of the Restricted Inverse Gamma (IG) Distribution to Log Normal and Empirical Distribution</td>
<td>95</td>
</tr>
<tr>
<td>3. Summary of Descriptive Statistics</td>
<td>96</td>
</tr>
<tr>
<td>4. Mixture Model Parameter Estimates for Interpurchase Time</td>
<td>97</td>
</tr>
<tr>
<td>5. Hierarchical Bayes Parameter Estimates for Interpurchase Time</td>
<td>98</td>
</tr>
<tr>
<td>6. Parameter Estimates for Monetary Value</td>
<td>99</td>
</tr>
<tr>
<td>7. Comparison of Individual Interpurchase Time Estimates</td>
<td>100</td>
</tr>
<tr>
<td>8. Predictive Performance for Interpurchase Times (days)</td>
<td>101</td>
</tr>
<tr>
<td>9. Predictive Performance for Monetary Value ($)</td>
<td>102</td>
</tr>
<tr>
<td>10. Predictive Performance for Customer Valuation ($/day)</td>
<td>103</td>
</tr>
<tr>
<td>11. Predictions for the Top and Bottom 5 Customer Accounts (Sorted by Customer Valuation, Vj)</td>
<td>104</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
</table>
| 1. The Family of Inverse Gamma Distribution  
\((\alpha=4, \theta=0.001-0.002)\) | 105 |
| 2. Restricted and Unrestricted Densities | 106 |
| 3. Distribution of Purchase Occasions | 107 |
| 4. Distribution of Interpurchase Times \((\lambda_t)\)  
(Maximum Likelihood Estimates) | 108 |
| 5. Distribution of Average Purchase Amount \((\bar{y})\)  
(Maximum Likelihood Estimates) | 109 |
| 6. Relationship Between Interpurchase Time and Number of Purchases | 110 |
| 7. Relationship Between Purchase Amount and Number of Purchases | 111 |
| 8. Sequence of Gibbs Draws for \(\theta\) | 112 |
| 9. Posterior Distribution of Interpurchase Times  
Hierarchical Bayes Model | 113 |
| 10. Relationship Between Change in Interpurchase Time Estimates and Recency: Right Censored Spell Included | 114 |
| 11. Relationship Between Change in Interpurchase Time Estimates and Recency: Right Censored Spell Excluded | 115 |
| 12. Distribution of Customer Value \((V_i)\) | 116 |
CHAPTER I
INTRODUCTION

In direct marketing, practitioners often use the indexes of recency, frequency, and monetary value (RFM) to evaluate how important or valuable a customer is to their business (Stone 1994; David Shepard Associates 1995). Recency measures the time since the last purchase. Long periods of recency are widely believed to signal that the customer behavior has changed, possibly even to the point of not making any purchases in the future. Frequency is the number of purchase occasions during a given period of time, such as a month or a quarter, and identifies customers who are in more immediate contact with the firm. Monetary value is the average dollar amount of the purchases.

Based on these three criteria, a scoring assignment system is then used to calculate the importance or the value of each customer (Stone and Wyman 1986, p.141). Understanding this individual value could help direct marketing managers in several ways. For instance, it can be used to target customers more accurately; to improve the effectiveness of sales promotion; to develop forecasts and establish sales quotas and budgets; to increase salesforce performance and job satisfaction; and finally, it can also be used to provide an assessment of a firm's performance with potential profitability based on the marketing standpoint of view, besides the traditional financial
analysis (i.e., a more meaningful measure of direct marketing success). Therefore, the accuracy of these scoring systems is critical to the efficient identification of profitable customers and the success of direct marketing programs.

Unfortunately, the current method (scoring assignment system) utilized in direct marketing practice has some drawbacks. Assigning points subjectively is the most troublesome because it could be biased and misleading. Furthermore, the lack of an appropriate statistical model to specify the observations of RFM makes it very difficult to predict or to understand the impact of marketing or demographic variables on the customer interpurchase time. Therefore, a model that can both produce an integrated RFM index (i.e., the customer value) or quantity for each customer and incorporate the effect of exogenous variables (either marketing variables or demographic variables) is needed. The purpose of this dissertation is to develop a statistical model to achieve this research objective.

Reviewing the literature shows that in spite of the volume of academic research for investigating the interpurchase time have been done during the past three decades, the discrepancy between the outputs from models and the demand in practice is still large for several reasons. First, most of the previous research is focused on how to get a good estimation of a model's parameters. However, those parameters are not of primary interest to practitioners. In the parametric hazard model approach, for example, the model treats the individual average interpurchase time $\lambda_i$ as a nuisance parameter that is not required to be estimated. Therefore, these models are
investigating those coefficients of exogenous variables and the common parameters of heterogeneity. Since $\lambda_i$ is treated as a nuisance parameter, we cannot get the measurement of the individual customer value and consequently we cannot do any individualized communication or marketing activities.

Second, instead of modeling interpurchase time, some researchers proposed models that assume that the purchase frequency follows a Poisson process. Although the Poisson distribution (purchase frequency) and Exponential distribution (interpurchase time) are two sides of one coin, modeling the purchase frequency ignores the right censored interpurchase time that is not a very appropriate way to model the customer purchasing behavior because any information is valuable to the model estimation. Including the right censored data into the likelihood function is very important to the model estimation especially when there are few data points at the individual level that are fairly common phenomena in direct marketing. In addition, the Poisson purchase model also suffered the same problem as the interpurchase time model, i.e., it could only provide an aggregate level estimation result.

Finally, suppose we directly estimated the individual average interpurchase time $\lambda_i$ by the maximum likelihood method; the estimation would be very unstable due to the small sample size at the individual level. This problem is very similar to analyzing the scanner panel data. Traditionally, the researchers used to pool all individuals' information together to reduce the variation of estimation by either using the fixed effect
model, random effect model, semi-parametric model, or Bayesian model. Several researchers have shown that the hierarchical Bayes model outperforms all of the above models (Allenby, Jen, and Leone 1996). However, the hierarchical Bayes model has not been applied to the interpurchase time data analysis. The reason for this is that all of the previous hierarchical Bayes models were centered on the normal distribution family. That is, a normal-likelihood plus a normal-prior will get a normal-posterior. Since the interpurchase time is a positive random variable, we need to develop a hierarchical Bayes model that is built on a non-normal distribution system.

In this dissertation, therefore, a non-normal-distribution hierarchical Bayes model is developed to provide a direct answer for measuring the individual customer value. The key contributions of the proposed model are as follows:

1. The proposed model can generate estimates (i.e., the customer value) at the individual level. On the basis of this value, direct marketing managers can use the model to predict the probability of repeat purchase and the timing of the next purchase for every single customer in the company's database.

2. The model also incorporates each customer's characteristic variables which allows us to investigate the impact of these variables on the average interpurchase time.
3. This non-normal-distribution hierarchical Bayes model for the positive random variable can be applied to the other marketing model where a positive parameter is required to be estimated.

This dissertation is divided into five chapters. **Chapter II** includes a literature reviews of related models such as the proportional hazard rate model, the NBD model, and the Pareto-NBD model. **Chapter III** introduces the hierarchical Bayes model of interpurchase time with and without the demographic variables. Some simulation studies are used to evaluate the robustness properties of the proposed models and the model misspecification due to the integer restriction on the shape parameter of Gamma distribution. The performance of the estimated empirical distribution and the direct divergence based on the information theory for these models are examined. **Chapter IV** presents the empirical results based on a set of data from a real company's customer database to illustrate the proposed models and to compare the results of other models. Finally, **Chapter V** discusses the managerial implementation of the proposed models in direct marketing content and briefly sketches out future work.
CHAPTER II
REVIEW OF THE LITERATURE

Direct marketers routinely collect information about the timing and amount of customer purchases. In the case of business-to-business marketing, customers are individual firms who place orders through time. For consumer markets, customers are individual households that make frequent purchases from various product categories. In both cases, measures of recency and frequency are used to better understand the value of the customer.

The data analysis for the duration time has been applied extensively in many different disciplines for a very long time. For example, it is called lifetime data analysis (or reliability analysis) in the engineering field (Lawless 1982), survival data analysis in the biostatistics field (Crowley and Johnson 1982), and duration data analysis in the econometric field (Lancaster 1979; Heckman and Singer 1984; Butler and Worrall 1991). In our study, it is called the interpurchase time analysis. Compared to the other fields, a unique characteristic of the interpurchase time data in marketing is the nature of the panel data structure, which is assembled from a complete customer database.

The panel data analysis has become a more and more important and popular tool in marketing research in recent years (Guadagni and Little
1983; Krishnamurthi and Raj 1988). On the basis of the numerous literature regarding the models for the panel data, tremendous evidence shows that those models often provide many insights of marketing strategies and reliable results (Allenby and Lenk 1994; Rossi, McCulloch; and Allenby 1995).

Despite considerable research on panel data, it is surprising that the effort of the statistical model development for the business-to-business or consumer direct marketing is relatively indigent. In the following subsections, we review some relevant statistical models for the interpurchase time and purchase incidence models which are closely related to RFM measures.

### 2.1 NBD Model

The negative binomial distribution (NBD) models (Ehrenberg 1988; Morrison and Schmittlein 1988) examine purchase frequency measures in the random-effects framework with pool data across individual units of analysis. The NBD model is constructed by two assumptions: individual Poisson purchasing behavior and gamma heterogeneity of purchase rates. The first assumption is that the number of purchases made by customer $i$ in a period of length $[0,T]$ will be the Poisson distribution:

$$P(X_i = x | \eta_i) = \frac{\eta_i^x e^{-\eta_i}}{x!}$$  \hspace{1cm} (2.1)
where \( \eta_i = \frac{T_i}{\lambda_i} \) is the individual specific parameter of purchase rate and \( \lambda_i \) is the individual average interpurchase time. Because some customers have higher value of \( \eta_i \) while others have smaller value of \( \eta_i \), the assumption of this heterogeneity in purchase rates across customers is to allow \( \eta_i \) distributed as a gamma distribution with p.d.f.

\[
g(\eta_i|\alpha, \theta) = \frac{1}{\Gamma(\alpha)\theta^\alpha} \eta_i^{\alpha-1} e^{-\frac{\eta_i}{\theta}}, \quad \eta_i > 0
\]  

(2.2)

where \( \alpha \) and \( \theta \) are the gamma parameters. Note that when \( \alpha \) is an integer, the p.d.f. of (2.2) is called an Erlang-\( \alpha \) distribution. By combining (2.1) and (2.2), we can get the random-effect specification which aggregates across all customers to find the average probability of purchase for a random population member

\[
P(X_i = x | \alpha, \theta) = \int_0^\infty p(X_i = x | \eta_i, T) g(\eta_i|\alpha, \theta) d\eta_i
\]  

(2.3)

\[
= \left( \frac{x + \alpha - 1}{\alpha} \right) \left( \frac{\theta}{\theta + 1} \right)^x \left( \frac{1}{\theta + 1} \right)^\alpha
\]  

(2.4)

for \( x = 0, 1, \ldots \), and \( \alpha, \theta > 0 \), and \( \alpha \) integers. Since Ehrenberg (1959) first introduced this model to marketing literature, it has become one of the major marketing models that is widely applied to numerous situations (most of them are frequently purchased consumer products) during the past 30 years. In general, the model fits the distribution of a number of purchases very well at the aggregate level.
Under this aggregate model, however, the $\eta_i$ is treated as a nuisance parameter, which is the focal metric interested by direct marketing practitioners. In addition, modeling the number of purchases ($X_j$) means that we only have one observation per customer. That is, the original panel-type of data structure was collapsed to a data structure of which the pattern of interpurchase time was ignored. Therefore, when we use this parsimonious and appealing NBD model, we also lose a lot of valuable information. Finally, as Morrison and Schmittlein (1988) pointed out, incorporating the effect of marketing variables into the NBD model is probably the most important area for future research. To overcome these problems, modeling the interpurchase time instead of the number of purchases has been rapidly developed in marketing research.

### 2.2 Hazard Function Models

It has been suggested in recent marketing articles that proportional hazard regression can more effectively model the interpurchase times data. Helsen and Schmittlein (1993) examine both conventional and hazard rate models for household interpurchase times. Their results indicate that the proportional hazard regression methods outperform other procedures in terms of the stability and face validity of the estimates and in terms of predictive accuracy.

Let $T$ be the continuous random variable that represents the interpurchase time (or spell) of interest and let $h(t)$ denote the hazard
function of $T$. The hazard function is the conditional probability density function that a purchase is made at the end of a spell, given that no purchase has been made in that spell. It is defined as

$$h(t) = f(t|T > t) = \frac{f(t)}{\Pr(T > t)} = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)} \quad (2.5)$$

where $f(t)$ is the probability density function of $T$; $F(t)$ is the cumulative distribution function; and $S(t)$ is the survivor function. Following Cox (1972), the proportional hazards specification can be expressed as follows:

$$h(t|X\beta) = h_0(t) \cdot \Psi(X\beta) \quad (2.6)$$

where $h_0(t)$ denotes the baseline hazard and $\Psi(X\beta)$ is a function of the covariates. An important issue is incorporating the heterogeneity in the model. Basically, there are two ways to specify heterogeneity: (1) allowing for heterogeneity in the baseline hazard function $h_0(t)$; or (2) using a common baseline hazard function but allowing an intercept term $\beta_0i$ in $\Psi(X\beta)$ to vary across the population. In the later case, $\beta_0i$ has been modeled by fixed-effects specification (Gönül and Srinivasan 1993); random-effects specification (Banerjee and Bhattacharyya 1976; Gupta 1991; and Jain and Vilcassim 1991); or nonparametric specification (Heckman and Singer 1984). However, in the former case, $h_0(t)$ only has been investigated by random-effects specification (Schmittlein and Morrison 1983; Butler and Worrall 1991).

Unfortunately, the primary interest for direct marketing is the individual-level estimation of the average interpurchase time $\lambda_i$. As the
parameter of average purchase number $\eta_i$ in the NBD model, $\lambda_i$ is also
treated as a nuisance parameter. Although a fixed-effects specification will
yield an individual estimate, estimating a large number of parameters poses
a considerable computational burden.

### 2.3 Pareto-NBD Model

Schmittlein, Morrison, and Colombo (1987) have proposed a model to
integrate the information of purchase rate and retention rate. The SMC
model starts with two basic assumptions about purchasing behavior at the
individual level. First, it assumes that the number of purchases for
customer $i$ in the period $[0,T]$ has the Poisson distribution, $P[X_i = x | \eta_i]$, where
$\eta_i$ is the individual specific parameter of purchase rate. Secondly, the
duration time (say $t$ months) of customer $i$ with a company until s/he drops
out (i.e., is inactive) is assumed to be exponentially distributed: $\text{EXP}(t | \mu_i)$,
where $\mu_i$ is the individual specific parameter of retention rate (or dropout
rate). Therefore, each customer is characterized by two traits: a purchase
rate $\eta_i$ and retention rate $\mu_i$.

To capture the nature of heterogeneity, one of the approaches to do
this is the random effect model, which assumes that those heterogeneities in
purchase rates and retention rates across customers are independently and
identically distributed random variables. In the SMC model, the $\mu_i$'s and $\eta_i$'s
are assumed to follow different gamma distributions over customers, i.e.,
\[ \mu_i \sim \text{Gamma}(\gamma, \alpha) \]
and
\[ \eta_i \sim \text{Gamma}(s, \beta). \]

Then, the parameters of purchasing behavior functions \((\gamma, \alpha, s, \beta)\) are estimated by aggregating the individual information across the population.

Based on this model specification, the SMC model provides the desired inferences and predictions. First, the probability that a customer with a particular observed transaction history is still active at time \(T\) since trial, \(P_i[\text{Active}]\), is a function of \(\gamma, \alpha, s, \beta, X_i\) (number of purchases), \(t_i\) (duration time since last purchase, recency), and \(T\) (the duration time period of observation). Secondly, the model also can calculate the expected number of future purchases, \(X^*\), to be made in some future period of length \(T^*\), i.e.,
\[ \mathbb{E}[X^* | \gamma, \alpha, s, \beta, X_i, t_i, T, T^*]. \] (see Schmittlein and Peterson 1994 for a more detailed discussion).

The conceptualization and operationalization of customer value in the SMC model has its significant contribution. However, there are some shortcomings of the SMC model. First, this is still a random-effects model. Secondly, the model double counts the recency information because \(t_i\) is part of \(T\) but \(T\) already contributes to the likelihood through the Poisson distribution. Another drawback of the SMC model is that the model couldn't link the consumer's purchase rate and retention rate to the contents of marketing strategies or mix variables. So, simply based on the specification of the SMC model, it seems that the marketing manager could do nothing to increase the purchase rate or the retention rate unless we could integrate marketing variables into the model.
2.4 The Importance of Individual Estimates

In recent years, many companies have spent millions in creating their own consumer marketing databases. Unfortunately, most of these systems can only provide the information for the purpose of mass marketing, such as a mailing list. The effectiveness of this sort of direct marketing practice has been criticized because it fails to recognize the consumer as an individual with unique desires and preferences. And most importantly, it fails to recognize the value of each customer to the companies. As a result, the inefficient communication causes a great amount of wasted resources (costs) in direct marketing channels. The only solution for this problem is that we need to upgrade the customer database from the level of management information system (MIS) up to the decision support system (DSS).

The key to this solution of strategy development in database marketing (under DSS environment) is an individual customer lifetime value analysis (Holtz 1992; Nash 1993; and Hughes 1994). This customer value analysis can be applied to develop marketing tactics in many ways. For example, retailers can use this output to select customers for special promotions and create profiles to learn more about their customers. In addition to that, marketers can also connect this value to customers’ anniversaries, birthdays, or their stage of family life cycle, etc. Therefore, marketers can understand the needs and wealth of each individual customer and the impact on the "lifetime" value.
In sum, we can see that there is still a large potential area for developing the model in strategic database marketing. The model development can take two directions. First, improving the model itself. It includes relaxing model assumptions and new estimation methodologies. Second, we can extend the model by incorporating new variables into the model such as marketing variables and demographic variables. In this thesis, we propose a hierarchical Bayes model to calibrate customers' interpurchase time and purchase amount to achieve the above goal of model improvement. In addition, we also adopt the computationally efficient Gibbs-sampling method to estimate the model.
CHAPTER III
HIERARCHICAL BAYES MODEL

In this chapter we introduce a hierarchical Bayes random effects model of interpurchase times and monetary values, respectively, which will facilitate the in-depth study of customer valuation. We first present a basic hierarchical Bayes model for interpurchase time that incorporates the unobserved heterogeneity only. In this simple framework, we derive the conditional distributions of model parameters and show how to apply the Gibbs Sampling algorithm to generate unique individual estimates. In this section, we also show that the two-parameter inverse gamma distribution constitutes a rich family in which a wide variety of shapes of the distribution of a positive random variable can be generated by adjusting the value of the parameters.

Since there is a simplifying assumption needed to efficiently estimate the model, Section 2 will investigate the approximation ability of inverse gamma distribution to the other distribution of the positive random variable (e.g., Log-normal distribution) and robustness of the restricted condition on one of the parameters of the Inverse Gamma distribution. The simulation studies show that the restricted inverse gamma distribution performs very well.
On the basis of the simple hierarchical Bayes model that was developed in Section 1, we extend the model by considering covariates into the hierarchical Bayes model. Section 3 will introduce how to incorporate covariates (i.e., observed heterogeneity) into the model where these covariates serve to shift the scale parameter of the inverse gamma distribution. In Section 4, we will present the methodology to perform the hierarchical Bayes model for the monetary value and compare this model to the traditional empirical Bayes model.

3.1 HB Model for Interpurchase Time

Suppose that the i-th customer has $n_i$ purchase incidents in the time period which spans $[0,T_i]$, let $(t_{i1}, t_{i2}, \ldots, t_{in_i})$ denote the corresponding interpurchase times, then $\sum_{j=1}^{n_i} t_{ij} = T_i$ and $T_i \leq T$ because each customer's first purchase starts at a different calendar time, in contrast to the laboratory experiment setting where every subject starts at the same time (i.e., $T_i = T$, for $i=1, \ldots, N$).

The model assumes that interpurchase times, $t_{ij}$, represent a random sample selected from an exponential distribution with the parameter, $\lambda_i$, where $\lambda_i$ is the average interpurchase time for the $i$-th customer and varies across customers of analysis according to a continuous distribution. The unit of $\lambda_i$ will be days per purchase in our empirical analysis, which we will discuss in the next chapter. The underlying probability density function (pdf) of $t_{ij}$ will depend on whether it is a censored or uncensored data. If
the i-th customer made the last purchase at time T, then $t_{i\text{in}_i}$ is an uncensored observation. Therefore, the pdf of $t_{ij}$ is:

$$f(t_{ij}|\lambda_i) = \frac{1}{\lambda_i} \cdot e^{-\frac{t_{ij}}{\lambda_i}} \quad \text{for } j = 1, 2, \ldots, n_i.$$ 

On the other hand, if the i-th customer made the last purchase before time T, then $t_{i\text{in}_i}$ is a censored observation and its pdf is:

$$S(t_{ij}|\lambda_i) = e^{-\frac{t_{ij}}{\lambda_i}} \quad \text{for } j = n_i.$$ 

As a result, the likelihood function of $\lambda_i$ for the i-th customer, $\ell(t_{ij}, \delta_{ij}|\lambda_i)$, is

$$\ell(t_{ij}, \delta_{ij}|\lambda_i) = \prod_{j=1}^{n_i} f(t_{ij}|\lambda_i)^{\delta_{ij}} \cdot S(t_{ij}|\lambda_i)^{1-\delta_{ij}}$$

where:

$$\delta_{ij} = \begin{cases} 1 & \text{if the observation is uncensored} \\ 0 & \text{if the observation is censored} \end{cases}$$

Substituting $f(t_{ij}|\lambda_i)$ and $S(t_{ij}|\lambda_i)$ into the likelihood function, we can get

$$\ell(t_{ij}, \delta_{ij}|\lambda_i) = \prod_{j=1}^{n_i} \lambda_i^{-\delta_{ij}} \cdot e^{-\frac{t_{ij}}{\lambda_i}} = \lambda_i^{-\Sigma_{j=1}^{n_i} \delta_{ij}} \cdot e^{-\Sigma_{j=1}^{n_i} \frac{t_{ij}}{\lambda_i}} \quad (3.1)$$

We allow for heterogeneity across customers by modeling the exponential parameter, $\lambda_i$, to be distributed across customers with an Inverse Gamma distribution with shape and scale parameters $\alpha$ and $\theta$. 
\[ \pi(\lambda_i | \alpha, \theta) = \frac{1}{\Gamma(\alpha) \theta^\alpha \lambda_i^{\alpha+1}} \cdot e^{-\frac{1}{\theta \lambda_i}} \]  

(3.2)

with mean \( \frac{1}{(\alpha - 1) \theta} \), variance \( \frac{1}{(\alpha - 1)^2 (\alpha - 2) \theta^2} \) and \( \alpha, \theta > 0 \).

In contrast to our model, other models of purchase incidence have often assumed a poisson distribution for the number of purchases \( z \) in a given time period \( T \), and a gamma distribution of heterogeneity of the purchase rate \( \lambda_i \) (Ehrenberg 1988; and Schmittlein, Morrison, and Columbo 1987). The purchase rate is the reciprocal of the interpurchase time (i.e. \( \lambda_i = 1/\lambda_i \)). In theory, there is no difference between a model, which assumes exponential interpurchase times with an inverse gamma distribution of heterogeneity, and one which assumes a poisson purchase rate with a gamma distribution of heterogeneity -- both will yield mathematically equivalent inferences and predictions. In practice, however, differences exist when it is not clear when a customer first becomes a customer (i.e., \( T_i < T \)).

As with any statistical model, a sample selection bias will exist if, in the poisson model, \( T \) is measured from the first purchase (i.e., the data is a function of the dependent variable). To avoid this bias, we elect to model the recency and frequency aspects of purchase behavior using the interpurchase time -- inverse gamma model specified by Equations (3.1) and (3.2).

The inverse gamma distribution, with diffuse priors on \( \alpha \) and \( \theta \), is treated as part of the prior specification in our Bayesian model. Figure 1
presents a set of inverse gamma distributions where the shape parameter is fixed at value 4, while the scale parameter vary from 0.001 to 0.002. As shown in Figure 1, this family of distributions is very flexible and produces tractable posterior distributions, which allow for the assessment of individual customer behavior. In Section 3.3, we will allow this scale parameter to be a function of covariates.

In general, Bayesian methods base assessment of model parameters $\Theta$ on the posterior distribution, defined as:

$$\pi(\Theta|\text{Data}) \propto \ell(\text{Data}|\Theta) \pi(\Theta)$$

For our problem, the posterior distribution is very complicated and contains $\alpha$, $\theta$, and $\{\lambda_i, i=1, ..., N\}$. This causes problems in assessing the posterior distribution of $\alpha$, $\theta$ or, any specific $\lambda_i$. For example, the posterior of $\lambda_i$ for the $i$-th customer is equal to:

$$\pi(\lambda_i|\text{Data}) = \int \cdots \int \pi(\lambda_i, \alpha, \theta|\text{Data}) \, d\alpha \, d\theta \, d\lambda_{-i}$$

where "-i" denotes "all units except i".

The evaluation of this integral has previously been unfeasible because of its high dimensions. However, recent advances in Monte-Carlo Markov chain estimation (e.g., Gibbs Sampling) now make this evaluation feasible. Gibbs Sampling is a simulation-based estimation procedure in which random draws are recursively simulated from the full conditional distributions of the model, and are used as conditioning arguments in subsequent draws. Upon convergence, these draws are from the true
posterior distribution (Gelfand and Smith, 1990). For our problem, the Gibbs chains generates draws from the following conditional distributions:

A. Conditional Distribution of $\lambda_i$:

The posterior distribution of $\lambda_i$, given the data $t_{ij}$, $\delta_{ij}$, and the other parameters, $\alpha$ and $\beta$, is

$$\pi[\lambda_i | t_{ij}, \delta_{ij}, \alpha, \beta] \propto \ell(t_{ij}, \delta_{ij} | \lambda_i) \pi(\lambda_i)$$

Substituting Equations (3.1) and (3.2) into the above equation, we can get

$$\pi[\lambda_i | t_{ij}, \delta_{ij}, \alpha, \beta] \propto \left( \prod_{j=1}^{n_i} \lambda_i^{-\delta_{ij}} e^{-t_{ij} / \lambda_i} \right) \left( \lambda_i^{-(\alpha+1)} e^{-1 / \theta \lambda_i} \right)$$

$$= \lambda_i^{-(\Sigma \delta_{ij} + \alpha + 1)} e^{-\frac{\sum t_{ij} + \theta^{-1}}{\lambda_i}}.$$

This formulation can be identified as proportional to the Inverse Gamma distribution, $\text{IG}(A, B)$, where

$$A = \alpha + \sum_{j=1}^{n_i} \delta_{ij}$$

and

$$B = \left( \theta^{-1} + \sum_{j=1}^{n_i} t_{ij} \right)^{-1} \text{ for } i=1,...,N. \quad (3.3)$$
Consequently, $\lambda_i$ is generated from this inverse gamma distribution and then the estimate of $\lambda_i$ will be the mean of draws after convergence.

**B. Conditional Distribution of $\theta$:**

The posterior distribution of $\theta$, given the other parameters, is

$$\pi(\theta|\lambda_i, \alpha) \propto \ell(\lambda_i|\alpha, \theta) \pi(\theta)$$

According to the Equation (3.2), the likelihood of $\theta$ is

$$\ell(\lambda_i|\alpha, \theta) = \prod_{i=1}^{N} \ell(\lambda_i|\alpha, \theta) = \prod_{i=1}^{N} \frac{1}{\Gamma(\alpha) \theta^\alpha \lambda_i^{\alpha+1}} \cdot e^{-\frac{1}{\theta \lambda_i}}$$

The natural conjugate prior distribution of $\theta$ is Inverse Gamma distribution, $\text{IG}(a_0, b_0)$. Therefore, the posterior distribution of $\theta$ is

$$\pi(\theta|\lambda_i, \alpha) \propto \left( \prod_{i=1}^{N} \frac{1}{\Gamma(\alpha) \theta^\alpha \lambda_i^{\alpha+1}} \cdot e^{\frac{-b_0}{\theta}} \right) \cdot \left( \frac{\theta^{\alpha N + a_0} e^{-\frac{b_0}{\theta}}}{\Gamma(a_0 + N)} \right)$$

Again, this formulation can be identified as proportional to the inverse gamma distribution, $\text{IG}(A, B)$, where

$$A = \alpha N + a_0$$
\[ B = \left( \frac{1}{b_0} + \sum_{i=1}^{N} \lambda_i^{-1} \right)^{-1} \]  

(3.4)

Consequently, \( \theta \) is generated from this Inverse Gamma distribution and then the estimate of \( \theta \) will be the mean of draws after convergence.

**C. Conditional Distribution of \( \alpha \):**

The posterior distribution of \( \alpha \), given the other parameters, is

\[ \pi[\alpha | \{ \lambda_i \}, \theta] \propto \ell[\{ \lambda_i \} | \alpha, \theta] \pi(\alpha) \]

According to the Equation (3.2), the likelihood of \( \alpha \) is

\[ \ell[\{ \lambda_i \} | \alpha, \theta] = \prod_{i=1}^{N} \ell(\lambda_i | \alpha, \theta) \]

\[ = \prod_{i=1}^{N} \left( \frac{1}{\theta \lambda_i} \right)^{\alpha - 1} e^{\theta \lambda_i} \]

\[ = \prod_{i=1}^{N} \frac{1}{\Gamma(\alpha) \lambda_i} \]

Unfortunately, there is no natural conjugate prior distribution of \( \alpha \). However, if \( \alpha \) is restricted to take on integer values by specifying a Discrete Uniform distribution as the prior distribution of \( \alpha \), then the posterior distribution becomes
\[ \pi(\alpha^*|\lambda_i, \theta) \propto \prod_{i=1}^{N} \left( \frac{1}{\theta \lambda_i} \right)^{\alpha^* - 1} e^{-\frac{1}{\theta \lambda_i}} \alpha^*! \]

\[ = \prod_{i=1}^{N} \Pr(\alpha^* = \alpha^* | \frac{1}{\theta \lambda_i}) \quad (3.5) \]

where \( \alpha^* = \alpha - 1 \) and \( \Pr(\alpha^* = \alpha^* | \frac{1}{\theta \lambda_i}) \) is the probability of the Poisson distribution, given the parameter \( 1/\theta \lambda_i \), \( \alpha^* = 1, 2, 3, \ldots, M \).

Consequently, in order to generate \( \alpha^* \), we first calculate the values of the discrete pdf of \( \alpha^* \) from the N joint Poisson distribution, then draw \( \alpha^* \) based on this distribution.

We summarize the Gibbs Sampling procedure for this Hierarchical Bayes model (Model I) as follows:

Step 1: Calculate the initial estimate of \( \lambda_i \), \( \hat{\lambda}_i \) based on the maximum likelihood method, i.e.,

\[ \hat{\lambda}_i = \frac{\sum_{j=1}^{n_i} t_{ij}}{\sum_{j=1}^{n_i} \delta_{ij}} \]

Step 2: Set the prior values of \( a_0, b_0, \) and \( M \). Generate initial values of \( \theta \) from IG(\( a_0, b_0 \)) and \( \alpha \) from Discrete Uniform(M).

Step 3: Generate a random draw of \( \lambda_i \) from the Inverse Gamma distribution of (3.3) based on the most current value of \( \alpha \) and \( \theta \).
Step 4: Generate a random draw of $\theta$ from the Inverse Gamma distribution of (3.4) based on the most current value of $\alpha$ and $\lambda_i$, $i=1,...,N$.

Step 5: Generate a random draw of $\alpha$ from the distribution of (3.5) based on the most current value of $\theta$ and $\lambda_i$, $i=1,...,N$.

Step 6: Iterate between steps 3 to 5 until convergence.

Step 7: Calculate the posterior means and standard deviations of $\alpha$, $\theta$, and $\lambda_i$, $i=1,...,N$.

We have written FORTRAN subroutines (see Appendix A) on the SUN-UNIX machine to estimate the individual interpurchase time, $\lambda_i$, and the heterogeneity parameters $\alpha$ and $\theta$.

The efficient implementation of the Gibbs sampling algorithm requires $\alpha$, the shape parameter of the inverse gamma distribution, to be restricted to take on integer values. This results in an estimation algorithm for which estimates are obtained quickly and easily, even for large datasets. We demonstrate in the next section that this restriction minimally affects the ability of inverse gamma distribution to represent heterogeneity in the exponential parameter $\lambda_i$.

**3.2 Robustness of the Restricted Inverse Gamma Distribution**

From the previous section we know that the distribution theory for our Bayesian model is very tractable if $\alpha$ is restricted to take on integer values. The purpose of this section is to investigate the effect of the integer restriction on the inverse gamma distribution's ability to reflect heterogeneity
for a variety of known distributions. To evaluate this restriction we selected a number of different positive value random variable distributions (e.g., Log-Normal Distribution and Inverse Gamma Distributions) as the true distribution, and measured the difference to the best approximating distribution with integer $\alpha$. Differences between the distributions are measured in terms of their moments and directed divergence (Kullback 1978), which is defined as:

$$I(1:2) = \int \left( \log \frac{f_2(x)}{f_1(x)} \right) f_1(x) \, dx$$

where $f_1(x)$ is the true distribution and $f_2(x)$ is the approximating distribution. The directed divergence provides a more comprehensive measure of difference between the distribution than that provided by the first few sample moments.

Table 1 reports the moments and directed divergence for three different inverse gamma distributions along with the best approximating distribution with $\alpha$ restricted to be integer. Figure 2 provides a graphical summary of the approximation. The solid line represents the unrestricted distributions and the dotted lines represent the restricted distributions. Parameters for the unrestricted distribution were chosen so that the restriction would have the greatest adverse effect. This was accomplished by setting the "true" value of $\alpha$ to take on values of $\hat{\alpha}$.5, where $\hat{\alpha}$ is varied.

The table and figure indicate that the inverse gamma distribution is robust to the integer restrictions on $\alpha$ for a wide variety of distributions. For example, the divergence measure indicates that there is only a .3%
difference between the restricted and unrestricted distributions when the true $a$ is 3.5. In addition, the reported moments and densities indicate a close approximation.

Table 2 reports the directed divergence measure for a variety of assumed distributions. The left column of the table reports parameters and summary statistics (mean, variance, skewness) of the assumed log normal distribution, the middle column reports the divergence and parameters of the best approximating unrestricted inverse gamma distribution, and the right-most column reports the divergence and parameters for the best approximating restricted inverse gamma distribution. In general, the directed divergence measures are extremely small. This demonstrates that the inverse gamma distribution is capable of approximating a variety of known distributions; and that the ability is relatively unhampered by the integer restriction on $a$.

We also make comparison to an unrestricted inverse gamma distribution with parameters derived from our data. This was accomplished by a 2-step process in which maximum likelihood estimates of interpurchase times were first obtained from each panel member in our data, and then these point estimates were treated as known values in obtaining the MLE for the inverse gamma parameters ($\alpha$ and $\theta$). The last row of the table reports the divergence of the inverse gamma distributions from an empirical distribution based on the data. The empirical distribution was obtained by using maximum likelihood estimates of interpurchase times for each panel member in our data (see next chapter for a full description). These inter-
Purchase time estimates for each panel member were assumed to be normally distributed with mean equal to the MLE and variance equal to the inverse of the observed Fisher information matrix. The normal distributions (one for each panel member) were then summed across the panel members to produce the "true" distribution reported in the left column of the table. Consistent with our finding with the log-normal distribution, we find that the integer restriction has little impact on the distribution's ability to represent heterogeneity.

3.3 Incorporating Covariates into the HB Model

Incorporating covariates into a marketing model is an essential task for strategy development in practice. The model incorporates covariates by equating the scale parameter of the inverse gamma distribution to predefined dichotomous variables: \( \theta_i = \exp(x'_i \beta) \)

where

\[
X_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ik} \\ \vdots \\ x_{im} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \\ \vdots \\ \beta_m \end{bmatrix} \quad \text{for } i=1, 2, ..., N,
\]

and \( x_{ik} \) takes on values of zero or one. That is, our model allows for the introduction of dummy variables, which serve to shift the scale parameter of the inverse gamma distribution. The heterogeneity distribution then becomes:
As shown in Figure 3.1, a smaller value of $\theta_i$ not only shifts the curve to the right but also inflates the curve that means the distribution of interpurchase time has a higher mean and variance (i.e., more heterogeneity). According to the equations (3.1) and (3.6), the Gibbs chains generates draws from the following conditional distributions:

**A. Conditional Distribution of $\lambda_i$:**

The posterior distribution of $\lambda_i$ given the data $t_{ij}$, $\delta_{ij}$, $X_i$ and the other parameters $\alpha$ and $\beta$ is

$$\pi(\lambda_i|t_{ij}, \delta_{ij}, \alpha, X_i, \beta) \propto \ell(t_{ij}, \delta_{ij}|\lambda_i) \pi(\lambda_i)$$

Substituting Equations (3.1) and (3.6) into the above equation, we can get

$$\pi(\lambda_i|t_{ij}, \delta_{ij}, \alpha, X_i, \beta) \propto \left( \prod_{j=1}^{n_i} \lambda_i^{\delta_{ij}} \cdot e^{\frac{-\delta_{ij}}{X_{ij}}} \right) \left( \lambda_i^{-(\alpha+1)} \cdot e^{\exp(\frac{\beta^T X_i)}{\lambda_i}}} \right)$$

$$= \lambda_i^{-(\alpha+\delta_{ij}+\alpha+1)} \cdot e^{\frac{-\delta_{ij} [\exp(\beta^T X_i)]^{-1}}{\lambda_i}}.$$

Again, this formulation can be identified as proportional to the

$$\text{Inverse Gamma} \left( \sum_{j=1}^{n_i} \delta_{ij} + \alpha, \left( \sum_{j=1}^{n_i} t_{ij} + [\exp(X_i)^{-1}]^{-1} \right)^{-1} \right) \quad (3.7)$$
B. Conditional Distribution of $\beta_k$:

The posterior distribution of $\beta_k$, given the other parameters, is

$$\pi(\beta_k | \lambda_i, \alpha, x_{ik}) \propto \prod_{i=1}^{n} \ell(\lambda_i | \alpha, \beta_k, x_{ik}) \pi(\beta_k)$$

In order to generate $\beta_k$, we first reparametrize $\beta_k$ by letting $e^{\beta_k} = \phi_k$ and then derive the posterior distribution of $\phi_k$. We also assume that all the covariates $X_k (k=1,2,...,m)$ are 0-1 dummy variables. Since

$$e^{X_i \beta} = e^{\sum_{k=1}^{m} x_{ik} \beta_k} = \prod_{k=1}^{m} (e^{\beta_k})^{x_{ik}} = \prod_{k=1}^{m} \phi_k^{x_{ik}}$$

we can rewrite the equation (3.6) as

$$\pi(\lambda_i; \alpha, \phi_k, x_{ik}) = \frac{1}{\Gamma(\alpha) \prod_{k=1}^{m} \phi_k^{x_{ik}} \lambda_i} \cdot e^{\left[ \sum_{k=1}^{m} \phi_k^{x_{ik}} \right]}$$

Define $\sum_{i=1}^{N} x_{ik} = N_k$ and $D^{(j)} = \prod_{k=1 \sim k=j}^{m} \phi_k^{x_{ik}}$. The likelihood of $\beta_k$ in terms of $\phi_k$ is:
\[ \ell(\phi_k | x_i, \lambda_i) \propto \left( \prod_{i=1}^{N} \phi_1^{-\alpha x_{i1}} \right) \left( \prod_{i=1}^{N} \phi_2^{-\alpha x_{i2}} \right) \cdots \left( \prod_{i=1}^{N} \phi_k^{-\alpha x_{ik}} \right) \cdots \left( \prod_{i=1}^{N} \phi_m^{-\alpha x_{im}} \right) \times \exp \left[ -\sum_{i=1}^{N} \left( \phi_1^{-x_{i1}} \cdot \phi_2^{-x_{i2}} \cdots \phi_k^{-x_{ik}} \cdots \phi_m^{-x_{im}} \cdot \lambda_i^{-1} \right) \right] \]

\[ \propto \phi_k^{-\alpha \sum_{i=1}^{N} x_{ik}} \cdot \exp \left[ -\phi_k^{-1} \sum_{i \in N_k} (\lambda_i D_i^{(k)})^{-1} + \sum_{i \notin N_k} (\lambda_i D_i^{(k)})^{-1} \right] \]

\[ \propto \phi_k^{-\alpha N_k} \cdot \exp \left[ -\phi_k^{-1} \sum_{i \in N_k} (\lambda_i D_i^{(k)})^{-1} \right] \]

Since the natural conjugate prior distribution of \( \phi_k \) is Inverse Gamma distribution, IG(\( a_{k0}, b_{k0} \)), the conditional posterior of \( \phi_k \) is proportional to

\[ \pi(\phi_k | \lambda_i, x_{ik}, D_i^{(k)}, a_{k0}, b_{k0}) \propto \phi_k^{-\alpha_{N_k} + a_{k0} + 1} \cdot \exp \left[ -\phi_k^{-1} \left( b_{k0}^{-1} + \sum_{i \in N_k} (\lambda_i D_i^{(k)})^{-1} \right) \right] \]

Therefore, \( \phi_k \) can be generated from

\[ \text{Inverse Gamma} \left( \alpha_{N_k} + a_{k0}, \left( b_{k0}^{-1} + \sum_{i \in N_k} (\lambda_i D_i^{(k)})^{-1} \right)^{-1} \right). \quad (3.8) \]

C. Conditional Distribution of \( \alpha \):

The posterior distribution of \( \alpha \), given the other parameters, is
Based on the equation (3.2), the likelihood of \( \alpha \) is:

\[
\ell(\alpha | \lambda_i, \theta_i) = \prod_{i=1}^{N} \frac{1}{\Gamma(\alpha) \theta_i^\alpha \lambda_i^{\alpha+1}} e^{\frac{-1}{\theta_i \lambda_i}}
\]

\[
= \prod_{i=1}^{N} \left( \frac{1}{\theta_i \lambda_i} \right)^\alpha e^{\frac{-1}{\theta_i \lambda_i}} \frac{1}{\Gamma(\alpha) \lambda_i}
\]

Let \( \alpha^* = \alpha - 1 \). If the prior distribution of \( \alpha^* \) is Discrete Uniform(\( M \)), then the posterior distribution of \( \alpha^* \), given \( \lambda_i \) and \( \theta_i \), is proportional to

\[
\prod_{i=1}^{N} \text{poisson}(y_{i, \lambda_i})
\]  

(3.9)

which is the product of \( N \) poisson distributions.

We summarize the Gibbs Sampling procedure for this Hierarchical Bayes model (Model II) as follows:

**Step 1:** Calculate the initial estimate of \( \lambda_i, \hat{\lambda}_i \) based on the maximum likelihood method.

**Step 2:** Set prior value of \( a_k, b_k \), and \( M \). Generate initial value of \( \phi_k \) from \( \text{IG}(a_k, b_k) \) and \( \alpha \) from Discrete Uniform(\( M \)).

**Step 3:** Generate a random draw of \( \lambda_i \) from Inverse Gamma distribution of (3.7) based on the most current value of \( \alpha \) and \( \theta_i \).
Step 4: Generate a random draw of $\phi_k$ from Inverse Gamma distribution of (3.8) based on the most current value of $\alpha$ and $\lambda_i$, $i=1,...,N$. Then calculate $\beta_k$ and $\theta_i$.

Step 5: Generate a random draw of $\alpha$ from the distribution of (3.9) based on the most current value of $\theta_i$, $\lambda_i$, $i=1,...,N$.

Step 6: Iterate between steps 3 to 5 until convergence.

Step 7: Calculate the posterior means and standard deviations of $\alpha$, $\theta_i$, and $\lambda_i$, $i=1,...,N$.

We have written FORTRAN subroutines (see Appendix B) on the SUN-UNIX machine to estimate the individual interpurchase time, $\lambda_i$, and the heterogeneity parameters, $\alpha$ and $\theta_i$.

3.4 HB Model for Monetary Value

Next we turn our attention to the monetary aspect of the RFM metric. An assumption which could greatly simplify our analysis is to assume that the monetary value of the purchase order is independent of the interpurchase time process (see also Schmittlein and Peterson 1994). As discussed below, this assumption is supported in our data, and allows us to build a separate model of monetary value in which the dollar value of a purchase by customer $i$ at time $t$ is modeled as:

$$y_{ij} \sim \text{Normal}(\mu_i, \sigma^2)$$

$$\mu_i \sim \text{Normal}(x_i\beta, \gamma^2)$$
This specification of monetary values serves to shrink estimates of the individual averages ($\mu_i$) toward the group mean $\bar{x}\beta$. Posterior estimates of $\mu_i$ are easy to evaluate when $\beta$, $\sigma^2$ and $\gamma^2$ are known because the resulting posterior is normal with mean equal to:

$$
\left(\frac{\sigma^2}{\sigma^2 + \gamma^2}\right)\bar{y}_i + \left(\frac{\gamma^2}{\sigma^2 + \gamma^2}\right)X_i\beta
$$

where $\bar{y}_i$ is the average dollar order for the $i$th firm. Empirical Bayes models take advantage of this conditional closed form solution by substituting point estimates for $\beta$, $\sigma^2$ and $\gamma^2$ derived from the data. As with the interpurchase time model discussed above, recent advances in estimation methodologies using Gibbs sampling now make it possible to avoid these simplifying assumptions, allowing one to obtain exact unconditional estimates from the posterior distribution. The Gibbs Sampler recursively generates draws from the following conditional distributions:

**A. Conditional Distribution of $\mu_i$ ($i=1,...,N$, one firm at a time):**

$$
\mu_i | y, X, \beta, \gamma^2, \sigma^2 \sim \text{Normal} \left( \mu^*, (\gamma^{-2} + n_i \sigma^{-2})^{-1} \right)
$$

(3.10)

where

$$
\mu^* = \left[ \gamma^{-2} + n_i \sigma^{-2} \right]^{-1} \left( \gamma^{-2} (X_i\beta) + n_i \sigma^{-2} \bar{y}_i \right)
$$

and

$$
\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}.
$$
B. Conditional Distribution of $\beta$:

$$
\beta | \mu_1, X, \gamma^2 \sim \text{Normal} \left( \beta^*, \left( V_{\beta}^{-1} + \gamma^{-2} \sum_{i} x_{ik}x_{ik}' \right)^{-1} \right)
$$

(3.11)

where: Prior $\beta \sim \text{Normal}(b_0, V_\beta)$,

$$
\beta^* = \left[ V_{\beta}^{-1} + \gamma^{-2} \sum_{i} x_{ik}x_{ik}' \right]^{-1} \left( V_{\beta}^{-1} b_0 + \gamma^{-2} \sum_{i} x_{ik}x_{ik}' \hat{\beta} \right)
$$

and

$$
\hat{\beta} = \left( \sum_{i} x_{ik}x_{ik}' \right)^{-1} \left( \sum_{i} x_{ik} \mu_i \right).
$$

C. Conditional Distribution of $\sigma^2$:

$$
\sigma^2 | y, \mu_i \sim \text{Inverted Chi-squared} \left( d_0 + \sum_i n_i, D_0 + \sum_i \sum_j (y_j - \mu_i)^2 \right)
$$

(3.12)

where: Prior $\sigma^2 \sim \text{Inverted Chi-squared} (d_0, D_0)$ and $n_i$ = number of observations of firm $i$.

D. Conditional Distribution of $\gamma^2$:

$$
\gamma^2 | y, \mu_i \sim \text{Inverted Chi-squared} \left( s_0 + N, S_0 + \sum_i (\mu_i - x_i \beta)^2 \right)
$$

(3.13)
where: Prior $\gamma^2 \sim \text{Inverted Chi-squared} \ (s_0, S_0)$ and $N =$ number of firms.

We summarize the Gibbs Sampling procedure for this Hierarchical Bayes model (Model III) as follows:

Step 1: Calculate the initial estimate of $\mu_i$ by using $\bar{y}_i$, and then calculate $\hat{\beta}$.

Step 2: Set initial value of $\sigma^2$, $\gamma^2$, and prior value of $(b_0, V_0)$, $(d_0, D_0)$, and $(s_0, S_0)$.

Step 3: Generate a random draw of $\mu_i$ from the Normal distribution of (3.10), based on the most current value of $\beta$, $\sigma^2$, and $\gamma^2$.

Step 4: Generate a random draw of $\beta$ from the Normal distribution of (3.11), based on the most current value of $\mu_i$ and $\gamma^2$.

Step 5: Generate a random draw of $\sigma^2$ from the Inverted Chi-squared distribution of (3.12), based on the most current value of $\mu_i$.

Step 6: Generate a random draw of $\gamma^2$ from the Inverted Chi-squared distribution of (3.13), based on the most current value of $\mu_i$ and $\beta$.

Step 7: Iterate between steps 3 to 6 until convergence.

Step 8: Calculate the posterior means and standard deviations of $\mu_i$ and $\beta$.

We have written FORTRAN subroutines (see Appendix C) on the SUN-UNIX machine to estimate the individual monetary value, $\mu_i$, and the hierarchical parameters $\beta$.

On the basis of the results of individual estimates of interpurchase time ($\lambda_i$) and monetary value ($\mu_i$), and the assumption of independence
between these two components, we can define $V_i$ as a measure of the dollars per day a customer is expected to generate, by

$$V_i = \frac{\bar{\mu}_i}{\lambda_i} \quad \text{for } i=1, 2, \ldots, N \quad (3.14)$$

where "~" denotes the posterior expectation.

In the next chapter, we present the empirical results based on a set of data from a real company's customer database to illustrate the proposed models and to compare the results to other models.
CHAPTER IV

AN EMPIRICAL STUDY OF THE HIERARCHICAL BAYES MODEL FOR DIRECT MARKETING DATABASE

We estimate the model with direct marketing data from a business-to-business context. The data is a subset of data originally studied by Schmittlein and Peterson (1994) who employed a traditional random-effect model, which does not produce unique individual-level estimates. In this chapter, we first describe the nature of data to illustrate some specific issues related to a typical industrial customer database, followed by the empirical analysis results for our Model I, II, and III. Section 3 presents the comparison of the hierarchical Bayes formulation to a finite mixture model and also to an aggregate-level analysis similar to that obtained from a traditional random-effects model. At the end, some managerial implications for recency, frequency, and monetary components of the RFM measure are discussed in Section 4.

4.1 A Direct Marketing Customer Database

The data we used in this study is from a company that directly sells a variety of office products to business customers. As most direct marketing companies, this firm registers detailed information on each transaction made by each customer. Each record includes order date and dollar volume per
order, which are two key quantities for our analysis. An important figure of this database is that the purchase records are kept in a "window" that spans three years for some practical reasons (e.g., computer storage space limitation). In this case, therefore, short historical data will always be a problem for estimation no matter how long the customer databases have been built.

There are 4795 random samples selected from the database provided by Schmittlein and Peterson (1994). In order to calculate the interpurchase time, customers were included in our analysis if they had at least two purchases over a three-year period. Figure 3 plots the frequency distribution of the purchase occasions for the 3386 customers in the dataset, for whom at least 2 purchases were observed. The figure indicates that there are many customers with very limited purchase histories with which to draw inferences about their behavior. For example, nearly 65% of the customers in the study had 5 or fewer purchase occasions. This illustrates the problem encountered in many business-to-business direct marketing environments, where infrequent purchases are common and simple statistical procedure would yield unreliable estimates of customer recency and frequency.

Figure 4 depicts the frequency distribution of the individual average interpurchase time ($\lambda_i$) based on maximum likelihood estimates from Equation (3.1). Although these estimates consider the right-censored spell, they do not pool information across customers in the analysis. The dispersion of estimated interpurchase times is extremely wide, with 25% of
the customers purchasing at least every 4 months and 25% of the customers exceeding 473 days between purchases. The nature of this data is typical for this context. That is, in business-to-business marketing, a company’s customer database usually has a large number of customers but relatively short purchase histories for those customers. This is particularly true for growing businesses that bring in new customers quite regularly or for many new companies with short company histories. By looking at the shape of distribution, an inverse gamma distribution used in the proposed model seems very reasonable to fit this type of heterogeneity.

Figure 5 sketches the frequency distribution of the average purchase amount ($\bar{y}_i$) for each customer based on maximum likelihood estimates from Equation (3.10). Unlike the distribution of interpurchase time, which has a very long and significant tail on the right, this distribution is approximately normally distributed (regardless of the light right-tail). Therefore, using a normal distribution to model the heterogeneity of purchase amount across customers is a reasonable prior specification.

Table 3 summarizes the descriptive statistics for the number of purchase occasions, the average interpurchase time, and the average purchase amount in a 3-year time period. For this firm the median number of purchases is four, the median interpurchase time is about eight months, and the average of order size is approximately $105.

Next we turn our attention to the relationship among these three variables. First, we investigate the relationship between interpurchase time and the number of purchases by graphing the boxplot as shown in Figure 6.
A negative relationship is clearly observed because both the mean and variance of interpurchase times decrease when the number of purchases increases. Also note that the customers in the upper quartile of interpurchase time distribution (see Figure 4) have an average purchase history of only 2.24 purchases. As shown below, scoring systems based on recency, frequency, and monetary value tend to be inaccurate when purchase records are considered on an individual-by-individual basis and information is not pooled across customers.

Secondly, in contrast to the significant relationship between interpurchase time and the number of purchase occasions, the purchase amount (i.e., the average order size) did not relate to the number of purchases made by customers. Figure 7 shows that the mean and variance of purchase amount are about the same across different numbers of purchase occasions.

Finally and most importantly, the relationship between interpurchase time ($\lambda_i$) and the purchase amount ($\mu_i$) needs to be investigated because it is the key assumption we made in the previous chapter for calculating the customer valuation ($\mu_i/\lambda_i$). Let \( \left( \lambda_i, \bar{y}_i \right) \) be the MLE and \( \left( \tilde{\lambda}_i, \tilde{\mu}_i \right) \) be the HB estimates of $\lambda_i$ and $\mu_i$. In our analysis we find the correlation across customers between $\bar{\lambda}_i$ and $\bar{y}_i$ to be -0.031, $\lambda_i$ and $\tilde{\mu}_i$ to be -0.009, which lends support to our assumption of independence.
4.2 Analysis and Results

We first investigate the finite mixture model, which yields individual estimates of interpurchase time as a basis for comparison purpose (Kamakura and Russell 1989). Table 4 reports parameter estimates from this model in which the distribution of heterogeneity (Equation 3.2) is replaced with a simpler expression:

\[ \pi(\lambda_t) = \sum_k m_k \lambda_k \]

where \( k \) indexes the number of latent classes and the mass points, \( m \), sum to one. The best fitting model, measured in terms of Schwarz's (1978) Bayesian Information Criteria (BIC), contains six latent classes. The estimated inter-purchase times range from 7.9 days with a mass of 0.002 to 330.1 days with a mass of 0.654. The average interpurchase time estimate (\( \sum_k m_k \lambda_k \)) is equal to 253 days. The finite mixture model failed to locate point masses above 330 days because customers in the upper quartile of Figure 4 have so few purchase occasions. Conversely, there are many point masses with small interpurchase times because customers in the lower quartile of Figure 4 have many purchase occasions (12.56 orders on average).

Table 5 reports parameter estimates for two different proposed Hierarchical Bayes models. Model I does not include covariates while Model II includes covariates (dummy variables) to indicate customers in different SIC-specific firms: medical offices (98 firms), insurance-related (30 firms),
and attorneys (180 firms). Parameter estimates are based on 500 draws from the posterior distribution, after a 500 iteration "warm-up" of the Monte Carlo Markov chain. Time series plots of the draws indicate that draws from the Markov chain converge to the posterior distribution well before 500 iterations (see Figure 8). The average estimated interpurchase time from Model I is 252 days which closely agrees with the estimate from the finite mixture model. In addition, Model II indicates that insurance-related firms and medical offices have shorter interpurchase times than the other firms, indicating a greater opportunity for customer interaction and market feedback.

For direct marketing situations, recognizing these differences is important for several reasons. First, one can investigate whether certain customer types make small purchases with short cycles and others purchase larger amounts but with longer cycles. Second, a firm can use this information to better direct its sales force to the various types of customers to optimize both timing of the sales call and the amount of effort that should be spent in order to attempt to make the sale. This information can also be used to identify the best types of customers to solicit for new business.

Figure 9 provides graphs of the distribution of heterogeneity of the four types of customers. Insurance-related firms have the shortest interpurchase times and attorneys have the greatest interpurchase times. Figure 9 indicates that the posterior distribution of interpurchase times is more concentrated than the distribution of maximum likelihood estimates (Figure 4), but less concentrated than individual inter-purchase time
estimates from the mixture model. Individual estimates derived from the finite mixture model are a convex combination of the mass points, weighted by the posterior probabilities of membership to each point (see Kamakura and Russell 1989 for a detailed discussion). This implies that the individual estimates of $\lambda_i$ derived from the mixture model lie within the region (8-330) days. The Hierarchical Bayes estimates therefore offer a compromise between maximum likelihood estimates, which for many units of analysis are based on very few observations, and finite mixture model estimates, which tend to be overly restricted, particularly for customers who purchase infrequently.

Finally, we fit the monetary value data for Model III (Hierarchical Bayes Model) to complete the estimation analysis. The application of customer valuation will be discussed in detail in Section 4.4. The estimation results for the Empirical Bayes Model and Hierarchical Bayes Model is provided in Table 6. The average estimated purchase amount from the HB Model is $103.71, which closely agrees with the estimate from the empirical Bayes model ($104.82). In addition, the HB Model indicates that medical offices have higher purchase amounts per order than other firms. Although the estimation results are very similar in these two models, we will show in the next section that the HB Model still produces slightly more accurate predictions.
4.3 Model Comparison

We first show some illustrative customer estimates of interpurchase time. Table 7 provides a comparison of hierarchical Bayes, finite mixture and maximum likelihood estimates of $\lambda_i$ for 10 selected customers. The second column of the table reports the number of purchase occasions, while the third reports the sum of interpurchase times, including the right censored spell. The ratio of these two statistics is the maximum likelihood estimate, which seems to differ from both the finite mixture and hierarchical Bayes estimates reported in the two right-most columns. Both the finite mixture and the hierarchical Bayes models combine information from the likelihood with prior information. The mass points in the finite mixture model are considered as the prior, while the inverse gamma distribution in the hierarchical Bayes model is treated as part of the prior. These priors serve to pool information about interpurchase times ($\lambda_i$) across the customers in the analysis. The results are posterior estimates of a customer's interpurchase times, which move the MLE toward the average interpurchase time of 252 days (see Table 4). Further, the amount of shrinkage is related to the number of purchase occasions. A large number of purchase occasions results in relatively little shrinkage while a small number of purchase occasions leads to relatively more shrinkage toward the overall mean. In other words, as the number of purchase occasions increases, the posterior becomes more influenced by the likelihood (the MLE) and less by the prior (the overall mean). A limitation of the finite mixture
model is the fact that the posterior estimates of \( \lambda_i \) cannot exceed the mass point with the largest interpurchase time, resulting in a great deal of shrinkage of the estimated interpurchase time for those customers in the upper quartile of Figure 4.

In order to perform the predictive tests, we create a split sample by identifying those customers who had at least two purchase occasions in each of two time periods: March 1986 to September 1987 and October 1987 to March 1989. This resulted in 1394 customers for the analysis. We then re-estimated the models with the first half of the data (March 1986 to September 1987), and treated the second half of the data as a holdout sample for predictive testing. Table 8 reports the predictive performance of hierarchical Bayes, finite mixture, and maximum likelihood estimates. Reported is the mean absolute deviation (MAD) and root mean squared error (RMSE) of the observed interpurchase times in the second half of the data based on estimates (\( \lambda_i \)) from the first half of the data. Hierarchical Bayes predictions are based on the mean of the posterior distribution of \( \lambda_i \). The table indicates that the hierarchical Bayes model, which employs a continuous distribution of unobserved heterogeneity, results in superior predictions when compared to both the finite mixture model and the fixed-effect model which maximized the likelihood for each customer’s data individually.

Same predictive tests procedure was performed for the monetary value component. Table 9 reports results of the Empirical Bayes and Hierarchical Bayes procedures using the split-half data described above. The
results for the Bayesian models show that predictive gains are possible from shrinking the individual estimates of monetary value toward the grand mean. In addition, explicit consideration of the scale parameters of the normal distributions results in slightly more accurate predictions.

As a result, our analysis indicates that there exist substantial gains from pooling data across respondents for both the recency/frequency and monetary components of RFM scoring systems. We now examine these gains with respect to an overall score, which reflects the overall value of the customer.

The customer valuation defined by Equation (3.14) is calculated based on either MLE or hierarchical Bayes estimates of $\lambda_i$ and $\mu_i$. Table 10 reports on the predictive performance of the proposed scoring statistic, once again using the predictive datasets described above. Posterior estimates of $\lambda_i$ and $\mu_i$ were obtained from the first half of the data using the hierarchical Bayes model described earlier. These estimates were then used to predict the total dollar expenditure in the second half of the data. A similar set of calculations were performed using maximum likelihood estimates, which do not pool the data across respondents.

The top portion of the table indicates that scores based on the hierarchical Bayes model estimates for interpurchase times are substantially better than those based on maximum likelihood estimates. For example, the MAD decreases from 341 to 238 (30%) and the RMSE from 1389 to 443 (68%) when a hierarchical Bayes model of interpurchase times is used, which pools the data across customers rather than examining purchase
histories in isolation. However, in this empirical example, only marginal gains are realized by using hierarchical Bayes methods for the monetary value portion of the calculations. The reason is that there were a number of customers with very large dollar amounts and these values had a large influence on the posterior estimates. Trimming the sample to only include those customers with less extreme monetary values results in superior predictions for scores that employ hierarchical Bayes estimates for both expected interpurchase times and monetary values.

In the next section we use the model to better understand the relationship between recency and frequency in direct marketing and its managerial implications.

4.4 Integration of Recency and Frequency

Since the interpurchase time model incorporates the right-censored spell of the data in the likelihood, this offers a natural method for integrating the recency and frequency measures used to score individual customers. This is in contrast to typical direct marketing practice, which assigns an often arbitrary weight to each of these terms and then produces a summary score. An interesting question is whether the likelihood function correctly considers the recency component (the right-censored spell) of a customer's purchase history. If the right censored spell carries greater weight than that specified by the likelihood (indicating that the most recent data is more reliable or representative of the customer), then it may not be appropriate to
use an exponential distribution, which assumes a constant hazard rate to represent interpurchase times. This suggests the need for a model with a more dynamic representation of the purchase process.

We examine this issue by using the split sample dataset discussed above to evaluate predictive performance. We measure recency for a specific customer as the probability of observing no purchases in the right-censored spell of the data. This probability is calculated with the posterior distribution of the customer's expected interpurchase time, \( \pi(\lambda_i | \text{Data}) \). The resulting probability provides a measure of recency which can be compared across customers in our dataset. If the right-censored spell is long relative to the history of interpurchase times for the customer, then the resulting probability would be small, i.e., a small likelihood of observing no purchase occasions in the time period. We estimate the posterior distribution of \( \lambda_i \) for both halves of the data \( \{\lambda_{i1}, \lambda_{i2}\} \), \( i=1...N \) and compare changes to our probability measure. If the right censored spell is informative about interpurchase times beyond that which is captured by the likelihood, then changes in estimated interpurchase time should be related to the probability measure.

Figure 10 provides boxplots of the ratio estimated interpurchase times for each customer for the first and second halves (i.e. \( \lambda_{i2}/\lambda_{i1} \)). The figure indicates that while the expected interpurchase times in the second portion of the data are generally greater than that estimated in the first portion of the data, there is no apparent relationship to the probability measure. This indicates that recency is correctly incorporated into the analysis through the
likelihood. In contrast, Figure 11 displays boxplots of ratios of interpurchase times in which $\lambda_{i1}$ is calculated without the right-censored spell (we denote these estimates as $\lambda_{i0}$). This is based on the assumption that, in practice, the frequency measure does not reflect the right-censored spell; hence the need for recency. Figure 11 shows a direct relationship between the ratio $\lambda_{i2}/\lambda_{i0}$ and the probability measure, indicating that long periods of recency, which lead to low probabilities, should be used to revise the interpurchase times upwardly for the next period. This finding is consistent with the conventional wisdom in direct marketing.

Our analysis provides one possible explanation for the practice of considering both recency and frequency -- ignoring the right-censored spell results in biased estimates of interpurchase times, particularly when there exists limited purchase histories. We find that incorporating the right-censored spell through the likelihood leads to a measure which is independent of recency -- i.e., the likelihood provides the correct weighting of recency and frequency in our data.

**4.5 Managerial Implications**

One of the most commonly asked questions in business-to-business direct marketing is: How can a database system be used to target "best" opportunities? Bean (1995) states that although most companies recognize that not all customers provide equal value, few companies are able to identify which customers are "best" and which are least profitable. The key
characteristic of direct marketing is its ability to communicate to each
individual customer based on RFM and other demographic information
about each customer. The framework of customer valuation proposed in
this dissertation gives a direct answer to that question. Since the measure
of customer values in our model is expected dollar per day \( V_j = \mu_j / \lambda_j \), we can
forecast each customer's expected purchase amount in the future.

For example, if a marketing manager wants to predict the sales
volume for the next year, a simple approach is using \( V_j \) times 365 days then
summing them up across customers in the analysis or the entire database.
For this random sample data set, we calculate \( V_j \times 365 \) based on hierarchical
Bayes estimates of \( \mu_j \) and \( \lambda_j \) and plot the frequency distribution as shown in
Figure 12. The average and median of customer values are $193.09 and
$143.97, respectively. The advantages of evaluating customers by these
quantities are prominent.

The last column of Table 4.9 reports the expected dollar volume,
which indicates the value of a customer from top to bottom. For those who
are identified as the most profitable customers (the upper portion of Table
4.9), marketing managers should maintain a good relationship with them by
some special services such as personal contact, on-line ordering process,
higher financial support, etc. On the other hand, for those who are
recognized to be the least profitable customers (the bottom portion of Table
4.9), marketing managers should use different incentives such as coupons,
frequent buyer special programs, etc., to shorten the interpurchase time or
increase the purchase amount per order.
In addition, since these quantities represent the potential profitability of current customer database which did not report in most traditional financial statements, it certainly can help corporate finance people to make a better acquisition valuation decision. A direct comparison of two customers' database can be done by simply computing $\sum_i V_i$. Another interesting application is a dynamic analysis of the firm's performance by using the split database discussed in Section 4.4. As shown in Figure 10, when we observe that the expected interpurchase times in the second portion of the data ($\lambda_{12}$) are generally greater than that estimated in the first portion of the data ($\lambda_{11}$), this might give us a warning sign of a firm's performance because the customers are gradually increasing their interpurchase time.
CHAPTER V

CONCLUSION AND DISCUSSION OF FUTURE WORK

This dissertation introduces a hierarchical Bayes model for interpurchase times with observed and unobserved components of heterogeneity. Unobserved heterogeneity is captured through an inverse gamma distribution for interpurchase times, while observed heterogeneity is introduced by allowing the location parameter of the distribution to depend on dummy variable covariates. We demonstrate the advantage of using a continuous mixing distribution of heterogeneity relative to a discrete distribution, and show how the model can be efficiently estimated with the Gibbs sampler when the shape parameter of the inverse gamma distribution, \( \alpha \), is restricted to take on integer values. Our analysis indicates that the inverse gamma distribution is robust to this restriction over a wide range of shape and scale parameters.

The hierarchical Bayes models, when estimated with simulation-based methods such as the Gibbs sampler, are particularly well suited to the analysis of direct marketing problems because they yield individual-level estimates as a by-product of the estimation procedure. These individual estimates are not available from traditional random-effect models which, for example, integrate the purchase incidence parameter out of the likelihood.
Algorithms are presented for estimating the proposed exponential/inverse gamma models in a Bayesian random-effect framework.

In addition to this methodological contribution, this dissertation makes a substantive contribution to the literature in examining recency, frequency, and monetary-value measures routinely used by direct marketers to evaluate their customers. Our analysis of individual-level estimates indicates that recency and frequency measures can be integrated by means of an exponential likelihood for interpurchase times, which represent recency as the right-censored spell of the data. Furthermore, assuming independence between these measures and the monetary value of an order results in a simple measure of valuation (dollars per day), which can be used to identify a firm's most profitable customers. We demonstrate that the use of hierarchical Bayes models for interpurchase times and monetary value result in a customer valuation which is more accurate than one based on estimates which do not pool the data or employ a discrete distribution of heterogeneity. The inverse gamma prior results in less severe shrinkage than a prior based on a finite set of mass points, while allowing for information to be pooled across units of analysis.

While the empirical example focuses on a specific business-to-business situation, it should be clear that this example is a member of a family of marketing problems which involve individual customer valuation. For example, since the model provides individual estimates of interpurchase time, it can be used to assess whether a firm's recent purchases deviate from their earlier purchases. An example of this is provided above in calculating
the probability of observing no purchases in the right-censored spell (see Figure 10 and 11). This probability can be used to identify customers for whom it is no longer profitable to direct promotional material. Alternatively, careful study of those firms who increase the frequency of orders may lead to the identification of profitable opportunities.

There exists a number of possible areas of future research of our basic model. First, our assumption of exponential interpurchase times may not be appropriate in some direct marketing contexts and may need to be relaxed. The exponential model implies a constant hazard rate which may be overly restrictive in many circumstances. Violations of this assumption would require longer purchase histories than that present in the business-to-business problem analyzed in this dissertation. Retailers with frequent shopper programs, credit card companies and catalogers are examples of firms where interpurchase times are shorter and purchase histories may be sufficiently long for the estimation of more complicated models. However, an advantage of our model is that it could be selectively applied to the most recent portion of the customer's purchase history (e.g., the last 6 months), and still yield estimates which are more accurate than estimates from other models. Whether the selective application of our simple model results in more accurate estimates of recency and frequency than complicated models which fully utilize the purchase history is an open question.

Second, we allow the location parameter of the inverse gamma distribution to be related to dummy variables. While many covariates used to model observed heterogeneity are measured on a nominal scale, it would
certainly be desirable to allow for the location parameter to be directly related to continuous covariates such as age or income. Last, we assume that interpurchase times and monetary values are independent of each other. While this assumption is supported in our empirical example, less restrictive association would certainly be desirable.

In summary, our analysis of purchase histories in a business-to-business direct marketing setting suggests that complicated scoring systems to combine recency, frequency, and monetary values are not always warranted. Recency and frequency can be naturally combined into one measure by modeling recency as the right censored spell in an exponential interpurchase time model. Furthermore, our analysis indicates that the estimated interpurchase times are unrelated to monetary value, leading to a simple calculation for the valuation of customers. The resulting metric, dollars per day, has intuitive appeal and can be directly related to the profitability of various actions routinely considered by direct marketers.
LIST OF REFERENCES


-----------, A. C. Bemmaor, and Donald G. Morrison (1985), "Why Does the NBD Model Work? Robustness in Representing Product Purchase,


APPENDIX A

FORTRAN PROGRAM FOR THE HB MODEL I

******************************************************************************
* setup.f ... program to estimate purchase data for
* hierarchical interpurchase time model.
* ***************

The sample size is 3386. nhh=3386
Assuming the time frame is 1102 days. nt=1102

implicit real*8 (a-h,p-z)
parameter (nhh=3386,nt=1102,nb=4)
dimension nday(nhh),nbuy(nhh),aday(nhh)
real*8 lambda(nhh)
integer hhid(nhh),t(nhh,nt),d(nhh,nt),index(nhh,3)
character*60 envvar
common /lnsiz/lnsize
common /toler/tol
lnsize=80

call getenv('td',envvar)
open(unit=15,file=envvar,status='old')
call getenv('index',envvar)
open(unit=16,file=envvar,status='old')

c read in data: td(i,j) and index(i,j)

do 10 i=1,nhh
   read(16,1) (index(i,j),j=1,3)
1   format(1x,3i8)
10 continue
do 20 i=1,nhh
   read(15,2) hhid(i),(t(i,j),j=1,index(i,3))
   read(15,2) id,(d(i,j),j=1,index(i,3))
60
format(i4,200i4)
continue

calculate the summation of t(i,j) and d(i,j)

c
nobs=1
do 50 i=1,nhh
   nday(nobs)=0
   nbuy(nobs)=0
   aday(nobs)=0
   do 40 j=1,index(i,3)
      nday(nobs)=nday(nobs)+t(i,j)
      nbuy(nobs)=nbuy(nobs)+d(i,j)
   40 continue
   day=nday(nobs)
   buy=nbuy(nobs)
   aday(nobs)=day/buy
   lambda(nobs)=aday(nobs)
   nobs=nobs+1
endif
50 continue
nobs=nobs-1
write (6,*) nhh,nobs

c
call main gibbs routine
c
call jgibbs(hhid,nday,nbuy,aday,lambd,lambda,nobs)
c
cstop
end

**********************************************************************
*                                                                        *
c  *                     jgibbs.f                                      *
c  *                                                                        *
c  **************************

main routine to cycle through gibbs iterations for hierarchical interpurchase time model...
subroutines used: jlambda,ja,jb,jout,and nag
routines for generating gamma, exponential, uniform, and poisson draws.

subroutine jgibbs(hhid, nday, nbuy, aday, lambda, nhh)
implicit real*8 (a-h,p-z)
real*8 lambda(nhh)
dimension nday(nhh), nbuy(nhh), aday(nhh)
integer hhid(nhh)
common /lnsiz/lnsize
lnsize=80
itermax=1000

first set up initial values

alpha=10
beta=0.0002
n0=20
alpha0=100
beta0=10
iter=0
99 continue
iter=iter+1
if(mod(iter,100).eq.0) then
write(6,* ) iter
endif

generate alpha
call ja(nhh, lambda, alpha, beta, n0, iter)

generate beta
call jb(nhh, lambda, alpha, beta, alpha0, beta0)

generate lambda(i)
call jlambda(nday, nbuy, nhh, lambda, alpha, beta)
cycle complete
write out parameters
if (iter.gt.500) then
  call jout(hhid, nhh, iter, lambda, alpha, beta,
          aday, n0, alpha0, beta0)
endif
if(iter.lt.itermax) goto 99
end

***************
*
* ja.f
*
***************

routine to generate alpha given everything

subroutine ja(nhh, lambda, alpha, beta, n0, iter)
  implicit real*8 (a-h,p-z)
  real*8 lambda(nhh), lnp(100), pdf(100), cdf(100)
  character*60 envvar
  common /lnsize/lnsize
  lnsize=80
  call getenv('alphack', envvar)
  open(unit=20, file=envvar)

  a=0
  b=1
  n=n0
  tlambda=0
  tilambda=0
  deno=0
  do 10 i=1,nhh
    tlambda=tlambda+log(lambda(i))
    tilambda=tilambda+(1/lambda(i))
  10 continue
  calculate log(Pr(alpha=1)) as starting point
  lnp(1)=-1*(tilambda/beta+nhh*log(beta)+tlambda)
  pdf(1)=exp(lnp(1))
deno=deno+pdf(1)
c
calculate log(Pr(alpha=2,3,...)) recursively
c
do 20 i=2,n
   ai=i
   il=i-1
   lnp(i)=lnp(il)-n*\log(ai)-n*\log(beta)-tlambda
   pdf(i)=exp(lnp(i))
   deno=deno+pdf(i)
20 continue
if (deno.eq.0) goto 50
c
c re-scale density
c
pdf(1)=pdf(1)/deno
cdf(1)=pdf(1)

if(cdf(1).eq.1.0) then
   alpha=1
   goto 999
endif
do 30 i=2,n
   pdf(i)=pdf(i)/deno
   cdf(i)=cdf(i-1)+pdf(i)
30 continue
cdf0=g05daf(a,b)
do 40 i=1,n-1
   if(cdf0.gt.cdf(i).and.cdf0.le.cdf(i+1)) then
      alpha=i+1
      goto 999
   endif
40 continue
goto 999
50 xmax=lnp(1)
alpha=1
c call dgmpnt(lnp,n,1)
do 60 i=2,n
   if(lnp(i).gt.xmax) then
      xmax=lnp(i)
      alpha=i
end if
continue
continue
alpha=alpha+1
    if(iter.le.50) then
        write(20,75) iter,alpha,beta
    do 70 i=1,n
        write(20,80) lnp(i),pdf(i),cdf(i)
    70 continue
    endif
    continue
75 format(1x,'iter=',i4,1x,'alpha=',f10.5,1x,
            'beta=',f10.6)
80 format(1x,f15.4,5x,e20.4,5x,e20.4)
return
end

******************************************************************************

* *
* jb.f *
* *
******************************************************************************

* routine to generate alpha given everything

subroutine jb(nhh,lambda,alpha,beta,alpha0,beta0)
implicit real*8 (a-h,p-z)
real*8 lambda(nhh)
common /lnsiz/lnsize
lnsize=80
slambda=0
do 10 i=1,nhh
    slambda=slambda+l/lambda(i)
10 continue
    a=int(nhh*alpha+alpha0)
    b=1/(slambda+1/beta0)
    ifail=0
    beta=1/g05d gf(a,b,ifail)
return
end

******************************************************************************
c  *
  *
  jlambda.f
  *
  ************************************************************
  *
  routine to generate Lambda(i) given everything
  *
  subroutine jlambda(nday, nbuy, nhh, lambda, alpha, beta)
    implicit real*8 (a-h,p-z)
    real*8 lambda(nhh)
    dimension nday(nhh),nbuy(nhh)
    common /lnsiz/lnsize
    lnsize=80
    do 10 i=1,nhh
      a=int(nbuy(i)+alpha)
      b=1/(nday(i)+1/beta)
      ifail=0
      lambda(i)=1/g05dgf(a,b,ifail)
    10 continue
  return
  end
  *
  jout.f
  *
  ************************************************************
  *
  program to compute mean and std of alpha, beta, 
  and lambda(i)
  *
  subroutine jout(hhid, nhh, iter, lambda, alpha, beta, 
    aday, n0, alpha0, beta0)
    implicit real*8 (a-h,p-z)
    real*8 mse, mad, lambda(nhh)
    dimension day(nhh), rlambda(5000,500), sa(500), sb(500), 
      r1(5000), r2(5000)
    integer hhid(nhh)
  c  linesize on output
  character*60 envvar
  common /lnsiz/lnsize
lnsize=80

```fortran
  call getenv('lambda',envvar)
  open(unit=16,file=envvar)
  call getenv('alphabeta',envvar)
  open(unit=17,file=envvar)
  call getenv('output',envvar)
  open(unit=18,file=envvar)

  nobs=iter-500
  do 10 i=1,nhh
    rlambda(i,nobs)=lambda(i)
  10 continue

  sa(nobs)=alpha
  sb(nobs)=beta
  write(17,500) iter,alpha,beta
  500 format(1x,i5,2(1x,f15.6))
  if (nobs.eq.500) then
    do 15 i=1,nhh
      do 14 j=1,nobs
        write(16,501) hhid(i),rlambda(i,j)
    14 continue
    15 continue
  sa1=0
  sa2=0
  sb1=0
  sb2=0
  do 30 i=1,nobs
    sa1=sa1+sa(i)
    sa2=sa2+sa(i)**2
    sb1=sb1+sb(i)
    sb2=sb2+sb(i)**2
  30 continue
  do 20 j=1,nhh
    r1(j)=r1(j)+rlambda(j,i)
    r2(j)=r2(j)+rlambda(j,i)**2
  20 continue
  do 30 continue

  c
  c now calculate the mean and std.
  c
  x=nobs
  am=sa1/x
```
as=\left(\frac{sa2-x*am^2}{x-1}\right)^{0.5}

bm=sb1/x
bs=\left(\frac{sb2-x*bm^2}{x-1}\right)^{0.5}

\text{write}(18,502) \text{ alpha0,beta0}

502 \text{ format}(1x, 'The prior of Beta:alpha0=',f6.0,'x beta0=',f5.3)

\text{write}(18,503) am,as

503 \text{ format}(' The estimation of alpha: mean=',f15.8,'x std=',f15.8)

\text{write}(18,504) bm,bs

504 \text{ format}(' The estimation of beta: mean=',f15.8,'x std=',f15.8)

\text{write}(18,505)

505 \text{ format}(' The estimation of individual lambda:')

\text{write}(18,506)

506 \text{ format}(1x, ' HHID', ' Point est. ', ' mean ', ' x ' ' std ')

\text{endif}
\text{return}
\text{end}
FORTRAN PROGRAM FOR THE HB MODEL II
(with covariate)

The sample size nhh=1202, 1935, 3386, 4795
Assuming the time frame is 1102 days. nt=1102
3 types of firm plus constant term, nb=4

implicit real*8 (a-h,p-z)
parameter (nhh=1232,nt=1102,nb=4)
dimension aday(nhh),nbuy(nhh),aday(nhh),
     a0(nb),b0(nb),phi(nb)
real*8 lambda(nhh),theta(nhh)
integer hhid(nhh),t(nhh,nt),d(nhh,nt),
     index(nhh,3),x(nhh,nb),x(1(nhh,nb)
character*60 envvar

common /lnsiz/lnsize
common /toler/tol
lnsize=80

call getenv('td',envvar)
open(unit=15,file=envvar,status='old')
call getenv('index',envvar)
open(unit=16,file=envvar,status='old')
call getenv('demo',envvar)
open(unit=17,file=envvar,status='old')
read in data: td(i,j) and index(i,j)

do 10 i=1,nhh
    read(16,1) (index(i,j),j=1,3)
 1 format(1x,3i8)
10 continue

do 20 i=1,nhh
    read(15,2) hhid(i),(t(i,j),j=1,index(i,3))
    read(15,2) id,(d(i,j),j=1,index(i,3))
 2 format(i4,200i4)
20 continue

do 30 i=1,nhh
    read(17,3) (xl(i,j),j=1,nb)
 3 format(7(i3))
*3 format(4(i1,1x))
30 continue

do 35 i=1,nhh
    x(i,1)=1
    do 32 j=2,nb
        x(i,j)=xl(i,j)
        x(i,j)=0
    32 continue
35 continue

calculate the summation of t(i,j) and d(i,j)

do 50 i=1,nhh
    nday(i)=0
    nbuy(i)=0
    aday(i)=0
    do 40 j=1,index(i,3)
        nday(i)=nday(i)+t(i,j)
        nbuy(i)=nbuy(i)+d(i,j)
    40 continue
    day=nday(i)
    buy=nbuy(i)
    aday(i)=day/buy
    lambda(i)=aday(i)
50 continue

c
first setup initial values

alpha=10
n0=20
aO(1)=11
bO(1)=50
do 60 i=2,nb
   aO(i)=21
   bO(i)=0.05
60 continue
do 70 i=1,nb
   a=aO(i)
   b=bO(i)
   ifail=0
   phi(i)=1/g05dgf(a,b,ifail)
70 continue
do 90 i=1,nhh
   theta(i)=1
   do 80 j=1,nb
      if (x(i,j).eq.1) then
         theta(i)=theta(i)*phi(j)
      endif
80 continue
90 continue

call main gibbs routine
call jgibbs(hhid,nday,nbuy,aday,lambda,alpha,n0,x,t,d,index,x,xl,a0,b0,phi,theta,nb,nhh,nt)
stop
end
and poisson draws.

subroutine jgibbs(hhid, nday, nbuy, aday, lambda, alpha, n0, t, d, index, x, x1, a0, b0, phi, theta, nb, nhh, nt)
implicit real*8 (a-h, p-z)
dimension nday(nhh), nbuy(nhh), aday(nhh),
a0(nb), b0(nb), phi(nb)
real*8 lambda(nhh), theta(nhh)
integer hhid(nhh), t(nhh, nt), d(nhh, nt),
index(nhh, 3), x(nhh, nb), x1(nhh, nb)
common /lnsiz/lnsize
lnsize=80
itermax=1000
iter=0
continue
iter=iter+1
if (mod(iter, 100).eq.0) then
write(6,*) iter
endif
c
generate alpha
c
call jal(nhh, lambda, alpha, theta, n0)
c
c generate phi(k)
c
call jphi(nhh, lambda, alpha, phi, a0, b0, x, nb)
c
do 90 i=1, nhh
theta(i)=1
do 80 j=1, nb
if (x(i,j).eq.1) then
theta(i)=theta(i)*phi(j)
endif
80 continue
90 continue
c
c generate lambda(i)
c
call jlambda(nday, nbuy, nhh, lambda, alpha, theta)
cycle complete

write out parameters

if(iter.gt.500) then
    call jout1(hhid,nhh,nb,iter,lamba,alfa,phi,aday)
endif
if(iter.lt.itermax) goto 99
end

* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
* jal.f
* *
* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

routine to generate alpha given everything

subroutine jal(nhh,lamba,alfa,theta,n0)
implicit real*8 (a-h,p-z)
real*8 lambda(nhh),theta(nhh),lnp(100),
     pdf(100),cdf(100)
common /lnsiz/lnsize
lnsize=80
a=0
b=1
n=n0
tllog=0
tlinv=0
deno=0
do 10 i=1,nhh
   tllog=tllog+log(lambda(i)*theta(i))
   tlinv=tlinv+(1/(lambda(i)*theta(i)))
10 continue

calculate log(Pr(alpha=1)) as starting point

lnp(1)=-1*(tllog+tlinv)
pdf(1)=exp(lnp(1))
deno=deno+pdf(1)
calculate \( \log(Pr(\alpha=2,3,\ldots)) \) recursively

do 20 i=2,n
   ai=i
   il=i-l
   lnp(i)=lnp(il)-nhh*\log(ai)-tllog
   pdf(i)=\exp(lnp(i))
   deno=deno+pdf(i)
20  continue
   if (deno.eq.0) goto 50

c re-scale density
c
   pdf(1)=pdf(1)/deno
   cdf(1)=pdf(1)
   if(cdf(1).eq.1.0) then
      alpha=1
      goto 999
   endif
do 30 i=2,n
   pdf(i)=pdf(i)/deno
   cdf(i)=cdf(i-1)+pdf(i)
30  continue
   cdf0=g05daf(a,b)
do 40 i=1,n-1
   if(cdf0.gt.cdf(i).and.cdf0.le.cdf(i+1)) then
      alpha=i+1
      goto 999
   endif
40  continue
   goto 999
50  xmax=lnp(1)
   alpha=1
c call dgmpnt(lnp,n,1)
do 60 i=2,n
   if(lnp(i).gt.xmax) then
      xmax=lnp(i)
      alpha=i
   endif
60  continue
continue
alpha=alpha+1
return
end

 subroutine jphi(nhh,lambda,alpha,phi,a0,b0,x,nb)
 implicit real*8 (a-h,p-z)
 real*8 lambda(nhh),ld(100)
 dimension a0(nb),b0(nb),phi(nb)
 integer x(nhh,nb),n(100)
 common /lnsiz/lnsize
 lnsize=80
 slambda=0
 do 30 k=1,nb
  ld(k)=0
  n(k)=0
 do 20 i=1,nhh
  dik=1
  if (x(i,k).eq.1) then
   n(k)=n(k)+1
   do 10 j=1,nb
    dik=dik*(phi(j)**x(i,j))
  10 continue
  dik=dik/phi(k)
  ld(k)=ld(k)+1/(lambda(i)*dik)
  endif
 do 40 k=1,nb
  a=int(n(k)*alpha+a0(k))
  b=1/(ld(k)+1/b0(k))
  ifail=0
  phi(k)=1/g05dgf(a,b,ifail)
 do 40 continue
return
subroutine jlambda(nday, nbuy, nhh, lambda, alpha, theta)
    implicit real*8 (a-h,p-z)
    real*8 lambda(nhh), theta(nhh)
    dimension nday(nhh), nbuy(nhh)
    common /lnsiz/lnsize
    lnsize=80
    do 10 i=1,nhh
        a=int(nbuy(i)+alpha)
        b=1/(nday(i)+1/theta(i))
        ifail=0
        lambda(i)=1/g05dgf(a,b,ifail)
    10 continue
    return
end

program to compute mean and std of alpha, beta, and lambda(i)

subroutine joutl(hhid, nhh, nb, iter, lambda, alpha, phi, aday)
    implicit real*8 (a-h,p-z)
    real*8 mse, mad, lambda(nhh)
    dimension aday(nhh), rlambda(6000,500), sa(500), sb(500,7), r1(6000), r2(6000), xm(6000),
    phi(nb), st(500,7), sb1(7), sb2(7), st1(7), st2(7)
    integer hhid(nhh)
    character*60 envvar
    common /lnsiz/lnsize
lnsize=80
call getenv('alphabeta', envvar)
open(unit=17, file=envvar)
call getenv('final', envvar)
open(unit=18, file=envvar)
nobs=iter-500
do 10 i=1,nhh
   rlambda(i,nobs)=lambda(i)
10 continue
sa(nobs)=alpha
do 20 j=1,nb
   sb(nobs,j)=phi(j)
   st(nobs,j)=phi(1)*phi(j)
   sb1(j)=0
   sb2(j)=0
   st1(j)=0
   st2(j)=0
20 continue
write(17,500) iter, alpha, (sb(nobs,j), j=1,nb),
x (st(nobs,j), j=2,nb)
500 format(lx,i5,8(lx,f15.6))
if (nobs.eq.500) then
   sal=0
   sa2=0
   do 50 i=1,nobs
      sal=sal+sa(i)
      sa2=sa2+sa(i)**2
   do 30 j=1,nb
      sb1(j)=sb1(j)+sb(i,j)
      sb2(j)=sb2(j)+sb(i,j)**2
      st1(j)=st1(j)+st(i,j)
      st2(j)=st2(j)+st(i,j)**2
   30 continue
   do 40 j=1,nhh
      r1(j)=r1(j)+rlambda(j,i)
      r2(j)=r2(j)+rlambda(j,i)**2
40 continue
50 continue

c
c now calculate the mean and std of alpha and beta.
x=nobs
am=sal/x
as=((sa2-x*am**2)/(x-1))**.5
write(18,503) am,as
503 format(' The estimation of alpha: mean=',f15.8,'
x std=',f15.8)
do 60 j=1,nb
   bm=sbl(j)/x
   bs=((sb2(j)-x*bm**2)/(x-1))**.5
   write(18,504) j,bm,bs
504 format(' The estimation of phi(',i1,' 
x'): mean=',f15.8,' std=',f15.8)
60 continue
do 70 j=2,nb
   bm=stl(j)/x
   bs=((st2(j)-x*bm**2)/(x-1))**.5
   write(18,505) j-1,bm,bs
505 format(' The estimation of theta(',i1,' 
x'): mean=',f15.8,' std=',f15.8)
70 continue
c
now calculate the mean and std of lambda(i).
c
write(18,506)
506 format(' The estimation of individual lambda:')
write(18,507)
507 format(1x,' HHID',' Point est. ',',', ' mean ',
x ' std ',')
do 80 i=1,nhh
   rm(i)=r1(i)/x
   rs=((r2(i)-x*rm(i)**2)/(x-1))**.5
   write(18,508) hhid(i),aday(i),rm(i),rs
508 format(1x,i5,3(1x,f12.4))
80 continue
endif
return
end
APPENDIX C

FORTRAN PROGRAM FOR THE HB MODEL III
(Monetary Value)

C--  *****************************************************
C--  *
C--  *  msetup.f . .  program to setup HB model for
C--  *  monetary value
C--  *
C--  *****************************************************
C--

C-- main routine to setup gibbs sampling for the model
C

C        Yij = Iij*Ui + eij
C        Ui = Xi*b + vi
C        e~Normal(0,D)
C        v~Normal(0,s2)
C
 implicit real*8 (a-h,p-z)
 parameter (nhh=1394,nt=551,nb=4)
 dimension index1(nhh,3),index2(nhh,3),
 x           tma(nhh,120),tmb(nhh,90),ave1(nhh),ave2(nhh),
 x           ui(nhh),beta(nb),d(nhh)
 integer x(nhh,nb)
 character envvar*60
 common /lnsiz/lnsize
 common /tolor/tol
 lnsize=80
 call getenv('tma',envvar)
 open(unit=11,file=envvar)
 call getenv('tmb',envvar)
 open(unit=12,file=envvar)
 call getenv('indexa',envvar)
 open(unit=13,file=envvar)
 call getenv('indexb',envvar)
 open(unit=14,file=envvar)
 call getenv('demoab',envvar)
open(unit=15,file=envvar)
call getenv('mle',envvar)
open(unit=16,file=envvar)

C--
read in the data
C--

do 10 i=1,nhh
   read(13,1) (index1(i,j),j=1,3)
   read(14,1) (index2(i,j),j=1,3)
   format(1x,3i8)
10 continue

do 20 i=1,nhh
   read(11,2) i,(tma(i,j),j=1,index1(i,3))
   read(12,2) i,(tmb(i,j),j=1,index2(i,3))
   format(1x,i4,150f6.0)
   format(1x,i4,150f10.0)
20 continue

do 50 i=1,nhh
   t1=0.0
   t2=0.0
   do 30 j=1,index1(i,3)
      t1=t1+tma(i,j)
   continue
   avel(i)=t1/index1(i,3)
   do 40 j=1,index2(i,3)
      t2=t2+tmb(i,j)
   continue
   ave2(i)=t2/index2(i,3)
   write(16,3) i,avel(i),ave2(i)
   format(lx,i4,2f15.4)
50 continue

do 60 i=1,nhh
   read(15,4) (x(i,j),j=1,nb)
   format(4(i1,1x))
   x(i,2)=0
   x(i,3)=0
   x(i,4)=0
60 continue

c
c call main gibbs routine
c
call jgibbs(nhh, nb, ave2, tmb, x, index2, ui, beta, d)
stop
end

c
...............................
*                       JGIBBS
*...............................
main routine to cycle through gibbs iterations for
Hierarchical Bayes Model for the Monetary Value
subroutines used: dsweep, dcond, wishsim, and nag
routines for generating normal, uniform, and
beta draws.
subroutine jgibbs(nhh, nb, ave1, tma, x, index1, ui, beta, d)
implicit real*8 (a-h,p-z)
dimension index1(nhh,3), ave1(nhh), tma(nhh,120),
x ui(nhh), beta(nb), d(nhh), xx(4,4), xy(4),
x varb(4,4), betaprio(4), work(100)
integer x(nhh,nb)
common /lnsiz/lnsize
lnsize=80
itermax=1000
eps=1.d-13
ier=0

c
initial values
c
do 30 i=1,nb
do 20 j=1,nb
do 10 k=1,nhh
    xx(i,j)=xx(i,j)+x(k,i)*x(k,j)
10    continue
20    continue
30    continue
    call dcond(xx, nb, work, 0)
call dsweep(xx, nb, eps, ier)
call dcond(xx, nb, work, 1)
do 50 i=1,nb
do 40 j=1,nhh
    xy(i)=xy(i)+x(j,i)*ave1(j)
40     continue
50     continue
   do 70  i=1,nb
      do 60  j=1,nb
         beta(i)=beta(i)+xx(i,j)*xy(j)
   60     continue
70     continue
   sigma=5
   betaprio(1)=100
   betaprio(2)=0
   betaprio(3)=0
   betaprio(4)=0
   do 80  i=1,nhh
      d(i)=20
80     continue
   varb(1,1)=1000
   do 90  i=2,nb
      varb(i,i)=100
90     continue
   call dcond(varb,nb,work,0)
   call dsweep(varb,nb,eps,ier)
   call dcond(varb,nb,work,1)
   iter=0
99     continue
   iter=iter+1
   if(mod(iter,100).eq.0) then
      write(6,*) iter
   endif
   c
c   generate Ui: the average purchase amount for firm i
   c
c      call lu(nhh,nb,avel,index1,x,beta,d,sigma,ui)
   c
c   generate beta: the coefficients of demographic variables
   c
c      call lb(nhh,nb,x,xx,beta,ui,sigma,betaprio,varb)
   c
c   generate d
   c
c      call ld(nhh,tma,ui,index1,d)
generate sigma

call ls(nhh,nb,ui,x,beta,sigma)

cycle complete

write out parameters

if(iter.gt.500) then
    call jout(nhh,nb,iter,ui,avel,beta,d,sigma)
endif

if(iter.lt.itermax) goto 99
end

*****************************************************************************
* *
*       LU
*
*****************************************************************************

routine to generate U_i given everything

subroutine lu(nhh,nb,avel,indexl,x,beta,d,sigma,ui)
imPLICIT real*8 (a-h,p-z)
dimension avel(nhh),indexl(nhh,3),ui(nhh),
        beta(nb),d(nhh)
integer x(nhh,nb)
common /lnsiz/lnsize
lnsize=80
eps=1.d-13
ier=0

generate U_i one firm at a time

do 20 i=1,nhh
   umer=0
   do 10 j=1,nb
      umer=umer+x(i,j)*beta(j)
  10   continue
ustart=(umer/(sigma**2)+avel(i)*indexl(i,3)/d(i))/
x (1/(sigma**2)+indexl(i,3)/d(i))
uvar=(1/(1/(sigma**2)+index1(i,3)/d(i)))**0.5
ui(i)=g05ddf(ustart,uvar)

continue
return
end

*********************************************************************************************************************
*                                                                LB
*                                                                *
*********************************************************************************************************************

routine to generate beta(i) given everything
subroutine lb(nhh,nb,x,xx,beta,ui,sigma,
              betaprio,varb)
implicit real*8 (a-h,p-z)
dimension ui(nhh),beta(nb),xx(4,4),xy(4),
              varb(4,4),varu(4,4),varp(4,4),betabar(4),
              betamean(4),work(1000),betaprio(4)
integer x(nhh,nb)
common /lnsiz/lnsize
lnsize=80
eps=1.d-13
ier=0
do 20 i=1,nb
   xy(i)=0
      do 10 j=1,nhh
         xy(i)=xy(i)+x(j,i)*ui(j)
      10 continue
20 continue
do 40 i=1,nb
   beta(i)=0
      do 30 j=1,nb
         beta(i)=beta(i)+xx(i,j)*xy(j)
      varu(i,j)=(sigma**2)*xx(i,j)
30 continue
40 continue
call dcond(varu,nb,work,0)
call dsweep(varu,nb,eps,ier)
call dcond(varu,nb,work,1)
do 60 i=1,nb
  do 50 j=1,nb
    varp(i,j)=varb(i,j)+varu(i,j)
  50 continue
60 continue
call dcond(varp,nb,work,0)
call dsweep(varp,nb,eps,ier)
call dcond(varp,nb,work,1)
do 80 i=1,nb
  betas=0
  betap=0
  do 70 j=1,nb
    betas=betas+varu(i,j)*beta(j)
    betap=betap+varb(i,j)*betaprio(j)
  70 continue
  betabar(i)=betas+betap
80 continue
do 95 i=1,nb
  betamean(i)=0
  do 90 j=1,nb
    betamean(i)=betamean(i)+varp(i,j)*betabar(j)
  90 continue
95 continue
call reference vector for multivariate draw
ciworkr=(nb+1)*(nb+2)/2
calc g05eaf(betamean,nb,varp,nb,eps,work,
x                          iworkr,ifail)
calc g05ezf(beta,nb,work,iworkr,ifail)
returnend
c
**********************************************************************
*                                                                     
*                    LD                                              
**********************************************************************

routine to generate di given everything
subroutine ld(nhh,tma,ui,index1,d)
implicit real*8 (a-h,p-z)
parameter (ia=6)
dimension indexl(nhh,3),ui(nhh),tma(nhh,120),
x d(nhh),wv(ia),wl(ia,ia),w2(ia,ia),w3(ia,ia),
x w4(ia,ia),w5(ia,ia)
common /lnsiz/lnsize
lnsize=80
idim=1

c generate di one firm at a time

c v=2
vi=10
resi=0
df=0
do 20 i=1,nhh

resi=0
do 10 j=1,indexl(i,3)
     resi=resi+(tma(i,j)-ui(i))**2
10 continue
    df=df+indexl(i,3)
20 continue
    idfg=v+df
    g=resi+vi

c generate sigma from inverse wishart dist.
c
call wishinv(g,idim,idfg,sigma2,wv,w1,w2,w3,w4,w5)
c
c the output from wishinv subroutine is sigma-square
c
do 30 i=1,nhh
    d(i)=sigma2
30 continue
return
end

c ***************
c * LS
c *
**routine to generate sigma given everything**

```fortran
subroutine ls(nhh,nb,ui,x,beta,sigma)
implicit real*8 (a-h,p-z)
parameter (ia=6)
dimension ui(nhh),beta(nb),yhat(4000),wv(ia),
x w1(ia,ia),w2(ia,ia),w3(ia,ia),w4(ia,ia),w5(ia,ia)
integer x(nhh,nb)
common /lnsize/lnsize
lnsize=80
eps=1.d-13
ier=0
idim=1
v=2
vi=4
resi=0
do 20 i=1,nhh
   yhat(i)=0
   do 10 j=1,nb
      yhat(i)=yhat(i)+x(i,j)*beta(j)
   10 continue
   resi=resi+(yhat(i)-ui(i))**2
20 continue
idfg=v+nhh
g=resi+vi
c
c generate sigma from inverse wishart dist.
c
call wishinv(g,idim,idfg,sigma,wv,w1,w2,w3,w4,w5)
c
c the output from wishinv subroutine is sigma-square

c sigma=sigma**0.5
return
end
```

```
program to compute mean and std of $U_i$, and beta

subroutine jout(nhh, nb, iter, ui, avel, beta, d, sigma)
implicit real*8 (a-h,p-z)
dimension ui(nhh), beta(nb), d(nhh), avel(nhh),
x rui(4000,500), sb(500,4), sb1(4), sb2(4), x r1(4000), r2(4000), rm(4000)
character*60 envvar
common /lnsiz/ lnsize
lnsize=80
call getenv('beta', envvar)
open (unit=17, file=envvar)
call getenv('final', envvar)
open (unit=18, file=envvar)
nobs=iter-500
do 10 i=1, nhh
   rui(i,nobs)=ui(i)
10 continue
do 20 i=1, nb
   sb(nobs,i)=beta(i)
20 continue
write(17,500) iter, (beta(i), i=1, nb)
500 format(1x,i5,4 (lx,fl5.6))
if (nobs.eq.500) then
do 50 i=1, nobs
   do 30 j=1, nb
      sb1(j)=sb1(j)+sb(i,j)
      sb2(j)=sb2(j)+sb(i,j)**2
30 continue
   do 40 j=1, nhh
      r1(j)=r1(j)+rui(j,i)
      r2(j)=r2(j)+rui(j,i)**2
40 continue
50 continue
c
now calculate the mean and std of $U_i$ and beta
c
y=nobs
do 60 j=1, nb
   bm=sb1(j)/y
   bs=(sb2(j)-y*bm**2)/(y-1)**.5
60 write(18,504) j, bm, bs
The estimation of beta(',il,'x'): mean=',fl5.8,' std=',f15.8)
continue
write(18,506)
The estimation of individual mean purchase amount:
write(18,507)
format(lx,' HHID',' Point est. ',' mean ',' std ')
do 80 i=1,nhh
    rm(i)=r1(i)/y
    rs=((r2(i)-y*rm(i)**2)/(y-1))**.5
    write(18,508) i,avel(i),rm(i),rs
format(lx,i5,3(lx,fl2.4))
80 continue
endif
return
end

***********************************************************************
DCOND 5/15/74

PURPOSE
IMPROVE ACCURACY OF DSWEEP WHEN DIAGONAL ELEMENTS OF A ARE NOT OF THE SAME ORDER OF MAGNITUDE.

USAGE
CALL DCOND(A,N,S,0)
CALL DSWEEP(A,N,EPS,IER)
CALL DCOND(A,N,S,1)

ARGUMENTS
A - SYMMETRIC POSITIVE DEFINITE N BY N MATRIX STORED COLUMNWISE.
    (STORAGE MODE 0)
    REAL*8
N - NUMBER OF ROWS AND COLUMNS OF A. LENGTH OF S.
    INTEGER*4
S - WORK VECTOR OF LENGTH N.
    REAL*8
ISW - INTEGER SWITCH.
THE USAGE
CALL DCOND(A,N,S,0)
WILL RETURN A CORRELATION MATRIX IN A AND THE
STANDARD ERRORS IN
S WHEN A IS A VARIANCE-COVARIANCE MATRIX.
SUBROUTINE DCOND(A,N,S,ISW)
REAL*8 A(N,N), S(N)
IF(ISW.EQ.1) GO TO 20
DO 10 I=1,N
10 S(I)=DSQRT(A(I,I))
DO 30 I=1,N
DO 30 J=1,N
30 IF(S(I)*S(J).NE.0.D0)
&A(I,J)=A(I,J)/(S(I)*S(J))
RETURN
END

DSWEEP 10/4/72

PURPOSE
INVERT A POSITIVE DEFINITE MATRIX IN PLACE BY A
DIAGONAL SWEEP.

USAGE
CALL DSWEEP(A,N,EPS,IER)

ARGUMENTS
A - SYMMETRIC POSITIVE DEFINITE N BY N MATRIX
STORED COLUMNWISE (STORAGE MODE OF 0).
ON RETURN CONTAINS THE INVERSE OF A
STORED COLUMNWISE.
REAL*8
N - NUMBER OF ROWS AND COLUMNS OF A.
INTEGER
EPS - INPUT CONSTANT USED AS A RELATIVE TOLERANCE IN
TESTING FOR
DEGENERATE RANK. A REASONABLE VALUE FOR EPS
SUBROUTINE DSWEEP(A,N,EPS,IER)
IMPLICIT REAL*8 (A-H,O-Z)
REAL*8 A(n*n)
TOL=0.DO
DO 5 I=1,N
II=N*(I-1)+I
TEST=A(II)
5 IF(TEST.GT.TOL) TOL=TEST
TOL=TOL*EPS
IER=0
DO 50 K=1,N
KK=N*(K-1)+K
AKK=A(KK)
IF(AKK.GT.TOL) GO TO 20
DO 10 J=K,N
KJ=N*(J-1)+K
10 A(KJ)=0.DO
IF(K.EQ.1) GO TO 16
KLESS1=K-1
DO 15 I=1,KLESS1
IK=KK-I
A(IK) = 0. DO
IER = IER + 1
GO TO 50
D = 1. DO/AKK
DO 25 I = 1, N
DO 25 J = I, N
IF ((I .EQ. K). OR. (J .EQ. K)) GO TO 25
IJ = N*(J-1)+1
IF (I .LT. K) AIK = A(N*(K-1)+I)
IF (I .GT. K) AIK = A(N*(I-1)+K)
IF (K .LT. J) AKJ = A(N*(J-1)+K)
IF (K .GT. J) AKJ = -A(N*(K-1)+J)
A(IJ) = A(IJ) - AIK*AKJ*D
25 CONTINUE
DO 30 J = K, N
KJ = N*(J-1)+K
30 A(KJ) = A(KJ)*D
IF (K .EQ. 1) GO TO 36
KLESS1 = K-1
DO 35 I = 1, KLESS1
IK = KK-I
35 A(IK) = -A(IK)*D
36 A(KK) = D
50 CONTINUE
DO 55 I = 1, N
DO 55 J = I, N
IF (I .EQ. J) GO TO 55
IJ = N*(J-1)+I
JI = N*(I-1)+J
A(JI) = A(IJ)
55 CONTINUE
RETURN
END

C
C ***************************************************************
C *
C * WISHINV
C *
C ***************************************************************
C
routine to set up call to wishsim for inverse wishart

```
subroutine wishinv(hh,idim,idf,d,wv,w1,w2,w3,w4,w5)
  implicit real*8 (a-h,p-z)
  dimension hh(idim,idim),d(idim,idim),wv(idim),
            w1(idim,idim),w2(idim,idim),w3(idim,idim),
            w4(idim,idim),w5(idim,idim)
  common /lnsiz/lnsize
  lnsize=80
  ifail=0
  eps=1.d-13
  ier=0

invert hh

  call dcond(hh,idim,wv,0)
  call dsweep(hh,idim,eps,ier)
  call dcond(hh,idim,wv,1)

call nag routine and obtain cholesky decompositon: L

  call f01bxf(idim,hh,idim,wv,ifail)
  do 18 i=1,idim-1
    do 17 j=i+1,idim
      w1(j,i)=hh(i,j)
    w1(i,j)=0.0
  17    continue
  18    continue
    do 20 i=1,idim
      w1(i,i)=1.0/wv(i)
  20    continue
  call wishsim(idim,w1,idf,w2,w3,w4,w5,d)
return
end
```
<table>
<thead>
<tr>
<th>True Distribution</th>
<th>Restricted Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. IG(3.5, 0.008)</td>
<td>1. IG(4, 0.007116)</td>
</tr>
<tr>
<td>$\mu=50.00$</td>
<td>$\mu=46.84$</td>
</tr>
<tr>
<td>$\sigma^2=1666.67$</td>
<td>$\sigma^2=1097.12$</td>
</tr>
<tr>
<td>Skewness* = 4.7199</td>
<td>Skewness* = 4.8308</td>
</tr>
<tr>
<td>$I(1:2)=0.003572$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2. IG(6.5, 0.002)</th>
<th>2. IG(6, 0.002176)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu=90.91$</td>
<td>$\mu=91.91$</td>
</tr>
<tr>
<td>$\sigma^2=1836.55$</td>
<td>$\sigma^2=2111.94$</td>
</tr>
<tr>
<td>Skewness* = 2.5955</td>
<td>Skewness* = 2.8297</td>
</tr>
<tr>
<td>$I(1:2)=0.001482$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3. IG(12.5, 0.0007)</th>
<th>3. IG(13, 0.000674)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu=124.22$</td>
<td>$\mu=123.64$</td>
</tr>
<tr>
<td>$\sigma^2=1469.67$</td>
<td>$\sigma^2=1412.46$</td>
</tr>
<tr>
<td>Skewness* = 1.3981</td>
<td>Skewness* = 1.3587</td>
</tr>
<tr>
<td>$I(1:2)=0.000094$</td>
<td></td>
</tr>
</tbody>
</table>

Note: (*) Sample Skewness Measure.
TABLE 2
Goodness of Fit of the Restricted Inverse Gamma (IG) Distribution
to Log Normal and Empirical Distribution

<table>
<thead>
<tr>
<th>True Distribution</th>
<th>Estimated Distribution</th>
<th>Unrestricted IG(α,θ)</th>
<th>Restricted IG(α,θ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-Normal(4, 0.3)</td>
<td>IG(11.74, 0.001635)</td>
<td>μ=56.55</td>
<td>μ=56.49</td>
</tr>
<tr>
<td></td>
<td>IG(12, 0.001600)</td>
<td>σ²=328.01</td>
<td>σ²=315.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Skewness*=1.39</td>
<td>Skewness*=1.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ι(1:2)=0.00771</td>
<td>Ι(1:2)=0.00784</td>
</tr>
<tr>
<td>Log-Normal(4, 0.5)</td>
<td>IG(4.20, 0.004930)</td>
<td>μ=63.24</td>
<td>μ=64.54</td>
</tr>
<tr>
<td></td>
<td>IG(4, 0.005176)</td>
<td>σ²=1642.76</td>
<td>σ²=2059.79</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Skewness*=3.25</td>
<td>Skewness*=4.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ι(1:2)=0.02016</td>
<td>Ι(1:2)=0.02080</td>
</tr>
<tr>
<td>Log-Normal(4, 0.7)</td>
<td>IG(2.13, 0.010914)</td>
<td>μ=79.91</td>
<td>μ=84.95</td>
</tr>
<tr>
<td></td>
<td>IG(2, 0.011623)</td>
<td>σ²=17520.66</td>
<td>σ²=23649.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Skewness*=15.6</td>
<td>Skewness*=17.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ι(1:2)=0.03062</td>
<td>Ι(1:2)=0.03172</td>
</tr>
<tr>
<td>Empirical</td>
<td>IG(1.31, 0.004810)</td>
<td>μ=634.71</td>
<td>μ=1175.77</td>
</tr>
<tr>
<td>Distribution</td>
<td>IG(1, 0.006301)</td>
<td>σ²=23518701</td>
<td>σ²=207899410</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Skewness*=51.1</td>
<td>Skewness*=58.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ι(1:2)=0.25352</td>
<td>Ι(1:2)=0.27487</td>
</tr>
</tbody>
</table>

Note: (*) Sample Skewness Measure.
### TABLE 3
Summary of Descriptive Statistics

<table>
<thead>
<tr>
<th>Descriptive Statistics</th>
<th>Number of Orders</th>
<th>Interpurchase Time ($\lambda_i$)</th>
<th>Average Purchase Amount ($\bar{y}_i$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Mean</td>
<td>5.97</td>
<td>343.65</td>
<td>104.82</td>
</tr>
<tr>
<td>2. Median</td>
<td>4</td>
<td>248.71</td>
<td>89.33</td>
</tr>
<tr>
<td>3. Mode</td>
<td>2</td>
<td>217.00</td>
<td>74.00</td>
</tr>
<tr>
<td>4. Std. Deviation</td>
<td>6.47</td>
<td>281.14</td>
<td>75.21</td>
</tr>
<tr>
<td>5. Minimum</td>
<td>2</td>
<td>6.54</td>
<td>7.00</td>
</tr>
<tr>
<td>6. Lower Quartile</td>
<td>2</td>
<td>134.38</td>
<td>66.17</td>
</tr>
<tr>
<td>7. Upper Quartile</td>
<td>7</td>
<td>473.00</td>
<td>121.00</td>
</tr>
<tr>
<td>8. Maximum</td>
<td>119</td>
<td>1095.00</td>
<td>1652.27</td>
</tr>
</tbody>
</table>
TABLE 4
Mixture Model Parameter Estimates for Interpurchase Time
(standard errors)

<table>
<thead>
<tr>
<th>Support Points ($\lambda_i$)</th>
<th>$m_i^*$</th>
<th>Point Mass $\lambda_i = \frac{e^{m_i^<em>}}{\sum_j e_j^</em>}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.9174 (0.2290)</td>
<td>0.0000</td>
<td>0.001859</td>
</tr>
<tr>
<td>19.7360 (0.7080)</td>
<td>0.8754</td>
<td>0.004462</td>
</tr>
<tr>
<td>38.0430 (2.5622)</td>
<td>2.1366</td>
<td>0.015749</td>
</tr>
<tr>
<td>69.2664 (4.4033)</td>
<td>3.8974</td>
<td>0.091614</td>
</tr>
<tr>
<td>129.0176 (9.5148)</td>
<td>4.8249</td>
<td>0.231617</td>
</tr>
<tr>
<td>330.0893 (10.2219)</td>
<td>5.8640</td>
<td>0.654698</td>
</tr>
</tbody>
</table>


TABLE 5
Hierarchical Bayes Parameter Estimates for Interpurchase Time
(posterior standard deviation)

<table>
<thead>
<tr>
<th>Parameters:</th>
<th>Model I (no covariates)</th>
<th>Model II (with covariates)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.00198123 (0.00002528)</td>
<td>0.00192991 (0.00002410)</td>
</tr>
<tr>
<td>$\theta_0$ (All Other Firms)</td>
<td></td>
<td>0.00311383 (0.00022377)</td>
</tr>
<tr>
<td>$\theta_1$ (Insurance-related Firms)</td>
<td></td>
<td>0.00223839 (0.00011930)</td>
</tr>
<tr>
<td>$\theta_2$ (Medical Offices)</td>
<td></td>
<td>0.00192939 (0.00041288)</td>
</tr>
<tr>
<td>$\theta_3$ (Attorneys)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Means (Days):</td>
<td>252.37</td>
<td>259.08</td>
</tr>
<tr>
<td>$1/(\alpha-1)\beta$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$1/(\alpha-1)\theta_0$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$1/(\alpha-1)\theta_1$</td>
<td></td>
<td>160.57</td>
</tr>
<tr>
<td>$1/(\alpha-1)\theta_2$</td>
<td></td>
<td>223.37</td>
</tr>
<tr>
<td>$1/(\alpha-1)\theta_3$</td>
<td></td>
<td>259.15</td>
</tr>
<tr>
<td>Parameters:</td>
<td>Empirical Bayes Model</td>
<td>Hierarchical Bayes Model</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>( \mu^* )</td>
<td>104.82</td>
<td>103.71</td>
</tr>
<tr>
<td>( \beta_0 ) (All Other Firms)</td>
<td></td>
<td>103.21</td>
</tr>
<tr>
<td>( \beta_1 ) (Insurance-related Firms)</td>
<td></td>
<td>0.4643</td>
</tr>
<tr>
<td>( \beta_2 ) (Medical Offices)</td>
<td></td>
<td>15.8658</td>
</tr>
<tr>
<td>( \beta_3 ) (Attorneys)</td>
<td></td>
<td>-0.4245</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>14255.02</td>
<td>13998.20</td>
</tr>
<tr>
<td>( \gamma^2 )</td>
<td>5656.01</td>
<td>2874.08</td>
</tr>
</tbody>
</table>
TABLE 7
Comparison of Individual Interpurchase Time Estimates

<table>
<thead>
<tr>
<th>Firm No.</th>
<th>Purchase Occasions</th>
<th>( \sum_{j=1}^{n_i} f_j )</th>
<th>MLE</th>
<th>Mixture Model</th>
<th>HB Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>2</td>
<td>787</td>
<td>787.00</td>
<td>325.73</td>
<td>429.67</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>687</td>
<td>687.00</td>
<td>323.16</td>
<td>406.68</td>
</tr>
<tr>
<td>19</td>
<td>3</td>
<td>1011</td>
<td>505.50</td>
<td>326.22</td>
<td>385.63</td>
</tr>
<tr>
<td>55</td>
<td>3</td>
<td>938</td>
<td>469.00</td>
<td>324.67</td>
<td>354.04</td>
</tr>
<tr>
<td>187</td>
<td>4</td>
<td>1066</td>
<td>355.33</td>
<td>322.59</td>
<td>321.89</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>970</td>
<td>323.33</td>
<td>318.51</td>
<td>287.01</td>
</tr>
<tr>
<td>45</td>
<td>6</td>
<td>1011</td>
<td>202.20</td>
<td>279.92</td>
<td>217.35</td>
</tr>
<tr>
<td>94</td>
<td>8</td>
<td>654</td>
<td>93.43</td>
<td>124.95</td>
<td>129.18</td>
</tr>
<tr>
<td>189</td>
<td>10</td>
<td>905</td>
<td>100.56</td>
<td>124.10</td>
<td>127.47</td>
</tr>
<tr>
<td>303</td>
<td>16</td>
<td>1008</td>
<td>67.20</td>
<td>78.50</td>
<td>89.18</td>
</tr>
</tbody>
</table>
TABLE 8  
Predictive Performance for Interpurchase Times (days)  

<table>
<thead>
<tr>
<th></th>
<th>MLE</th>
<th>Mixture Model</th>
<th>HB Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAD</td>
<td>109.7996</td>
<td>92.0889</td>
<td>81.8607</td>
</tr>
<tr>
<td>RMSE</td>
<td>149.1890</td>
<td>110.8794</td>
<td>101.5077</td>
</tr>
<tr>
<td></td>
<td>MLE</td>
<td>EB Model</td>
<td>HB Model</td>
</tr>
<tr>
<td>---------------</td>
<td>-------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>MAD</td>
<td>64.28</td>
<td>61.72</td>
<td>61.19</td>
</tr>
<tr>
<td>RMSE</td>
<td>100.31</td>
<td>98.49</td>
<td>97.74</td>
</tr>
</tbody>
</table>

TABLE 9
Predictive Performance for Monetary Value ($)

<table>
<thead>
<tr>
<th>Sample Size=1394</th>
<th>Customer Values</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MLE($)/MLE(λ)</td>
<td>HB($)/MLE(λ)</td>
<td>MLE($)/HB(λ)</td>
<td>HB($)/HB(λ)</td>
</tr>
<tr>
<td>1. MAD</td>
<td>341.00</td>
<td>285.16</td>
<td>238.51</td>
<td>230.99</td>
</tr>
<tr>
<td>2. RMSE</td>
<td>1388.64</td>
<td>765.92</td>
<td>443.23</td>
<td>457.54</td>
</tr>
<tr>
<td>Sample Size=1365*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. MAD</td>
<td>323.28</td>
<td>261.86</td>
<td>221.60</td>
<td>206.82</td>
</tr>
<tr>
<td>2. RMSE</td>
<td>1361.54</td>
<td>699.68</td>
<td>375.72</td>
<td>339.23</td>
</tr>
<tr>
<td>Sample Size=1330**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. MAD</td>
<td>304.44</td>
<td>241.49</td>
<td>200.39</td>
<td>179.57</td>
</tr>
<tr>
<td>2. RMSE</td>
<td>1356.38</td>
<td>679.33</td>
<td>332.16</td>
<td>268.45</td>
</tr>
</tbody>
</table>

Note: (*) Trimmed Sample Size: Delete the firms whose purchase amount at second time period is greater than 2500 or less than 50.

(**) Trimmed Sample Size: Delete the firms whose purchase amount at second time period is greater than 1500 or less than 50.
### TABLE 11
Predictions for the Top and Bottom 5 Customer Accounts
(Sorted by Customer Valuation, $V_j$)

<table>
<thead>
<tr>
<th>Firm No.</th>
<th>Purchase Occasions</th>
<th>Inter-purchase Time ($\lambda_i$)</th>
<th>Monetary Value ($\mu_i$)</th>
<th>Customer Valuation ($V_j$)</th>
<th>Expected $$ Volume for Next 1 Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>36</td>
<td>119</td>
<td>13.57</td>
<td>188.25</td>
<td>13.87</td>
<td>5062</td>
</tr>
<tr>
<td>3056</td>
<td>61</td>
<td>17.50</td>
<td>169.89</td>
<td>9.71</td>
<td>3544</td>
</tr>
<tr>
<td>673</td>
<td>11</td>
<td>131.24</td>
<td>1177.21</td>
<td>8.97</td>
<td>3274</td>
</tr>
<tr>
<td>1187</td>
<td>34</td>
<td>46.47</td>
<td>337.19</td>
<td>7.26</td>
<td>2648</td>
</tr>
<tr>
<td>356</td>
<td>17</td>
<td>83.08</td>
<td>422.87</td>
<td>5.09</td>
<td>1858</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>664</td>
<td>2</td>
<td>518.46</td>
<td>76.62</td>
<td>0.15</td>
<td>54</td>
</tr>
<tr>
<td>895</td>
<td>2</td>
<td>550.57</td>
<td>81.06</td>
<td>0.15</td>
<td>54</td>
</tr>
<tr>
<td>1929</td>
<td>2</td>
<td>539.80</td>
<td>79.14</td>
<td>0.15</td>
<td>54</td>
</tr>
<tr>
<td>1931</td>
<td>2</td>
<td>529.86</td>
<td>74.47</td>
<td>0.14</td>
<td>51</td>
</tr>
<tr>
<td>3345</td>
<td>2</td>
<td>539.30</td>
<td>74.00</td>
<td>0.14</td>
<td>50</td>
</tr>
</tbody>
</table>
FIGURE 1
The Family of Inverse Gamma Distribution
\((\alpha=4, \theta=0.001-0.002)\)
FIGURE 2
Restricted and Unrestricted Densities
FIGURE 3
Distribution of Purchase Occasions

Mean=6
Median=4
St.Dev.=6.5
Q1=2
Q3=7
N=3386
The Average Interpurchase Time (Days)

FIGURE 4
Distribution of Interpurchase Times ($\lambda$)
(Maximum Likelihood Estimates)

Mean = 344
Median = 249
St. Dev. = 281
Q1 = 134
Q3 = 472
N = 3386
FIGURE 5
Distribution of Average Purchase Amount \( \bar{y}_i \)
(Maximum Likelihood Estimates)

Mean = 104.82
Median = 89.33
St. Dev. = 75.21
Q1 = 66.17
Q3 = 121.00
N = 3386
FIGURE 6
Relationship Between Interpurchase Time and Number of Purchases
FIGURE 7
Relationship Between Purchase Amount and Number of Purchases
FIGURE 8
Sequence of Gibbs Draws for $\theta$
FIGURE 9
Posterior Distribution of Interpurchase Times
Hierarchical Bayes Model
FIGURE 10
Relationship Between Change in Interpurchase Time Estimates and Recency: Right Censored Spell Included
FIGURE 11
Relationship Between Change in Interpurchase Time Estimates and Recency: Right Censored Spell Excluded
FIGURE 12
Distribution of Customer Value ($V_j$)

Mean = 194.00
Median = 143.98
St. Dev. = 197.80
Q1 = 101.95
Q3 = 217.27
N = 3386