Marketing Letters 14:1, 5–20, 2003
© 2003 Kluwer Academic Publishers. Manufactured in The Netherlands.

A Bayesian Approach to Modeling Purchase Frequency

LICHUNG JEN lichung@mba.ntu.edu.tw
Department of International Business, National Taiwan University, 50, Lane 144, Sec. 4, Keelung Rd., Taipei, Taiwan

CHIEN-HENG CHOU chouchdr@ms36.hinet.net
Department of International Trade, Chinese Culture University, 55, Hwa Kang Rd., Yang Ming Shan Taipei, Taiwan

GREG M. ALLENBY allenby.1@osu.edu
Department of Marketing, Ohio State University, 2100 Neil Avenue, Columbus, OH 43210, USA

Revised 13 May 2002; Accepted 20 September 2002

Abstract

Direct marketers are often faced with the task of ranking, or scoring individual customers in terms of their expected value to the firm. A critical element of their scoring systems is expected frequency of customer interaction. In this paper the authors develop a hierarchical Bayes model of purchase frequency that combines a Poisson likelihood with a gamma mixing distribution, where the mixing distribution is a function of covariates. The proposed model is evaluated with two direct marketing datasets, and is shown to provide improved estimates of purchase frequency, particularly for customers with short purchase histories or who have infrequent interaction with the firm.

Keywords: purchase frequency, poison regression, NBD model, hierarchical Bayes

The frequency that a customer interacts with a firm is one of the most important indicators of the customer's financial value to the firm. Customers who frequently interact with the firm are expected to generate a greater stream of revenue, and often have a longer expected life, than those who interact infrequently (Blattberg et al., 2001). In direct marketing, the frequency of interaction is an important component of customer scoring algorithms that are used to rank customers in terms of their overall value (Stone, 1994; David Shepard Associates, 1995). However, estimates of expected purchase frequency are often very noisy because customer interactions do not follow regular patterns within fixed periods. Orders are often placed at irregular time intervals, and as a result simple estimates of expected frequency based on the number of orders placed during a prescribed period (e.g., the last six months) are sometimes unreliable. In this paper we propose a Hierarchical Bayes (HB) approach to estimating expected purchase frequency that are more reliable than simple estimates without placing undue restrictions on their variability.

Fixed-effects model-based estimates of purchase frequency typically employ a Poisson distribution for the number of orders placed in a given time period. Heterogeneity in purchase frequency can be incorporated into the Poisson model by relating the intensity parameter (λ) to observed covariates. A drawback of this approach is that customers with equal covariate values (e.g., all large buyers) are predicted to have the same purchase frequency in the next non-overlapping period.

Random-effects specifications avoid this problem by introducing unobserved heterogeneity into the model specification. The gamma distribution is often used as the random-effects distribution for the intensity parameter in the Poisson model, resulting in the well-known negative binomial distribution (Cameron and Trivedi, 1986; Ehrenberg, 1988). While the negative binomial model allows for more flexible dispersion of the intensity parameter, it does not provide individual-level estimates of purchase frequency. Although covariates can be incorporated into the resulting negative binomial likelihood, this model suffers from the same problem as the Poisson model – that is, it produces common estimates for customers with equal covariate values.

An alternative approach to developing individual-level estimates of purchase frequency is to employ the conditional NBD model (Morrison and Schmittlein, 1988). In this empirical Bayes model, the gamma random-effects distribution is combined with individual-level Poisson data in the calibration sample to generate predictions of future purchase frequency. An advantage of this approach is that the Bayesian updating equations are linear in the observed purchases and therefore easy to implement. A disadvantage is that covariates cannot be incorporated into the analysis.

We propose a HB model of purchase frequency that combines a Poisson likelihood with a gamma distribution of heterogeneity. Our hierarchical Bayes model allows for dispersion in the intensity parameter similar to the negative binomial distribution while producing individual-level estimates of purchase frequency, similar to the conditional NBD model of Morrison and Schmittlein (1988). Moreover, our approach allows for the introduction of covariates into the model specification. We find that the covariates aid in predicting purchase frequency, particularly for customers that infrequently interact with the firm, or have short purchase histories. Thus, our approach combines the advantage of obtaining unique individual-level estimates present in the conditional NBD model with the ability to enter covariates in the model specification as in traditional fixed-effects models.

In section 1 we develop our HB model and compare it to other commonly used methods of estimating purchase frequency. We then apply the model to direct marketing data and compare its performance to traditional methods, followed by concluding remarks.

1. The Model

Direct marketers routinely collect information about the timing 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 individuals or households that make frequent purchases from various product categories. The Poisson distribution is the simplest approach to modeling the number of purchases within a given time period. This model has been used in a wide variety of applications ranging from studying patent applications (Hausman et al., 1984) and highway fatalities (Michener and Tighe, 1992) to non-durable purchases (Bucklin et al., 1998). Denote y_i as the number of times that the ith customer interacts with a firm in a given time period of length T. Assume that y_i is distributed according to a Poisson distribution:

$$\Pr(y_i|\lambda_i, T) = (\lambda_i T)^{y_i} \exp(-\lambda_i T)/y_i!, \tag{1}$$

where $\lambda_i = E(y_i)/T$ is the purchase frequency. Covariates can be introduced into this model by relating them to the purchase frequency through a log linear model: $\lambda_i = \exp(x_i'\beta)$, where x_i denotes the vector of the ith subject's characteristic variables and β is the corresponding coefficient vector of x_i. Without loss of generality, we suppress the notation for the time period, T, in the following discussion and interpret λ_i as the expected number of orders placed by the ith customer in a given period.

Although the Poisson regression likelihood offers a simple approach to modeling purchase frequency, the model is restrictive in that it assumes that the mean and variance of the data are equal. In practice, the variance of the data is often greater than the mean, resulting in inconsistent estimates of model parameters when the Poisson likelihood is used (Grogger and Carson, 1991; Hausman et al., 1984; Shaw, 1988). To accommodate this over-dispersion, one can assume a random-effects distribution for λ_i instead of a log linear model. This assumption results in a model capable of reflecting data with greater variation. When the random-effects are assumed to be distributed gamma, the result is the negative binomial distribution (NBD) model.

1.1. The NBD Model

The NBD model uses the gamma distribution to integrate out the individual λ_i's across the population of customers, resulting in a marginalized likelihood that is a function of the parameters of the gamma distribution (Ehrenberg, 1988; Morrison and Schmittlein, 1988), which can be shown as follows:

$$\Pr(y_i|\alpha, \theta) = \int p(y_i|\lambda_i) g(\lambda_i|\alpha, \theta)\, \mathrm{d}\lambda_i = \binom{y_i + \alpha - 1}{y_i} \left(\frac{\theta}{\theta + 1}\right)^{y_i} \left(\frac{1}{\theta + 1}\right)^{\theta}, \tag{2}$$

where $g(\lambda_i|\alpha, \theta)$ is a gamma distribution, α is the shape parameter and θ is the location-scale parameter. The expected value of λ_i is equal to $\alpha\theta$. The likelihood of the NBD model is defined as the product of the term on the left hand side over the subjects in the sample.

Since the NBD model was first introduced by Ehrenberg (1959) to marketing literature, it has been used in many studies ranging from predicting new product sales to modeling inventory control (Brockett et al., 1996; Ehrenberg, 1988; Massy et al., 1970; Morrison and Schmittlein, 1988). The focus of these studies is to describe the distribution of purchase frequency in particular markets. That is, the NBD model yields estimates of α and θ which describe the distribution of λ_i, and do not yield individual-level estimates of λ_i. This is a shortcoming of the NBD model.

In many direct marketing applications, the focus of analysis is on the expected behavior of individual customers. Customers are frequently ranked according to the likelihood of their responses to a particular offering made by the firm, or in terms of their expected contribution to the long-term profitability of the firm. Disaggregate estimates of expected purchase frequency for each customer is therefore desired. While it is possible to introduce covariates into the NBD likelihood in Equation (2) (e.g., $E[\lambda_i|\alpha, \theta_i] = \alpha\theta_i = \exp(x_i'\beta)$), the model suffers from the same problem identified with the Poisson likelihood. That is, customers with the same covariate values are assigned the same expected purchase frequency.

1.2. The Conditional NBD Model

The conditional NBD model is a Bayesian version of NBD model (Morrison and Schmittlein, 1988). It predicts the number of purchases (Y_2) made by a person in a period of duration (t_2) on the condition of the number of purchases (Y_1) made by this same person in the previous non-overlapping period of duration (t_1). Assuming the prior distribution of λ to be a gamma (α, θ) mixing distribution, the conditional NBD model combines Poisson likelihood with the gamma prior to produce an updated gamma $(\alpha + y, \theta + 1)$ posterior distribution. Thus, the mean of this posterior distribution for a particular customer's purchase frequency, λ, is a conditional expectation. For $Y_1 > 0$, the conditional expectation is:

$$E[Y_2|Y_1 = y > 0; \alpha, \theta, t_1, t_2] = \frac{(\alpha + y)t_2}{\theta + t_1}. \tag{3}$$

For $Y_1 = 0$, the conditional expectation is:

$$E[Y_2|Y_1 = 0; \alpha, \theta, \gamma, t_1, t_2] = \frac{(1-\gamma)\left(\frac{\alpha t_2}{\theta+t_1}\right)\left(\frac{\theta}{\theta+t_1}\right)^{\alpha}}{\gamma + (1-\gamma)\left(\frac{\theta}{\theta+t_1}\right)^{\alpha}}, \tag{4}$$

where γ is the proportion of individuals that have purchase frequency $\lambda = 0$ in the dataset. To produce the conditional NBD estimates which not only consider individual purchase information but also covariates, one can use the estimates from the NBD model with covariates (i.e., $\theta_i = \exp(x_i'\beta)/\alpha$) as an input into Equations (3) and (4).

1.3. Hierarchical Bayes Model

The HB model combines the Poisson likelihood with the gamma random-effects distribution without integrating out the individual λ_i's. In contrast to Equation (2), the gamma distribution is viewed as part of the prior distribution to arrive at the posterior. In general notation, let Θ denote the parameters of a model. Then Bayes rule states that the posterior is proportional to the product of the prior and the likelihood:

$$\pi(\Theta|\text{Data}) \propto \ell(\text{Data}|\Theta)\pi(\Theta), \tag{5}$$

where ℓ denotes the likelihood, $\pi(\Theta)$ is the prior and $\pi(\Theta|\text{Data})$ is the posterior. A hierarchical Bayesian version of the NBD model then becomes:

$$\pi\left(\alpha, \theta, \{\lambda_i\}|\{y_i\}\right) \propto \prod_i p(y_i|\lambda_i)g(\lambda_i|\alpha, \theta)\pi(\alpha)\pi(\theta), \tag{6}$$

where $\pi(\alpha)$ and $\pi(\theta)$ are prior distributions on the parameters of the gamma distribution of λ_i, and the last three factors on the right hand side of the Equation (6) form the prior distribution. $\pi(\alpha)$ and $\pi(\theta)$ are typically specified to have minimal influence on the posterior distribution. Estimates of individual purchase frequency is obtained by integrating the joint posterior density:

$$\pi\left(\lambda_j|\{y_i\}\right) = \int \cdots \int \pi\left(\alpha, \theta, \{\lambda_i\}|\{y_i\}\right) \mathrm{d}\alpha\, \mathrm{d}\theta\, \mathrm{d}\lambda_{-j}, \tag{7}$$

where "$-j$" denotes "all units except j."

The result is individual level estimates of purchase frequency that reflect both the data from the individual customer (y_i) and other customers as expresses through the random-effects distribution. This combination of information results in individual estimates of purchase frequency that are less extreme than those based on the data alone. In addition, covariates can be introduced into the model specification by relating them to parameters of the gamma distribution: $\theta_i = \exp(x_i'\beta)$.

The evaluation of the integral in Equation (7) has previously been infeasible because of its high dimensions. However, recent advances in Markov chain Monte-Carlo estimation (e.g., Gibbs sampling) now make this evaluation feasible. Combining the Poisson likelihood with gamma heterogeneity is a relatively straightforward task. The introduction of covariates, however, severely complicates the Markov chain. In the appendix we provide an estimation algorithm in which θ, the location parameter of the gamma distribution, is related to dummy-variable covariates through a link function $\theta_i = \exp(x_i'\beta)$.

Markov chain Monte-Carlo estimation 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). Details of the computational algorithm are provided in the appendix. In the analysis presented in the next section, we estimate the model using 10,000 iterations of the Markov chain, from which the first 9500 iterations were discarded and the last 500 iterations were used to form estimates of the posterior distribution of model parameters. Time series plot of the draws indicate convergence of the chain from multiple initial values.

2. Empirical Applications

In this section we compare the in-sample and predictive fits of the hierarchical Bayes model to other methods of estimating expected purchase frequency with two datasets. The first dataset is from an office supply company engaged in business-to-business selling in the United States. The data span a period of three years, providing sufficiently long purchase

Table 1(a). Poisson, NBD, and Hierarchical Bayes Parameter Estimates for Purchase Frequency (Based on the Dataset of Business-to-Business Marketing Firm)

Parameters	Poisson Model (1)[f]	NBD Model (1)	Poisson Model (2)[g]	NBD Model (2)	HB Model
Intercept	1.6913693	1.7523893	1.1775850	1.1714120	0.7421252
	(0.048286)[e]	(0.065095)	(0.030287)	(0.044671)	(0.0469742)
Northeast	−0.0450466	−0.0474131	−0.0473115	−0.0434026	−0.0439138
	(0.017329)	(0.025599)	(0.017344)	(0.024860)	(0.0288897)
Midwest	−0.0496764	−0.0474233	−0.0461124	−0.0360934	−0.0362610
	(0.021605)	(0.035168)	(0.021599)	(0.034944)	(0.0399340)
South	−0.1031945	−0.1100203	−0.0836994	−0.0906082	−0.0903015
	(0.019271)	(0.028057)	(0.019296)	(0.028604)	(0.0354891)
Insurance-related	0.1390316	0.1429011	0.1550857	0.1345827	0.0820847
Firms	(0.071284)	(0.118839)	(0.071291)	(0.126180)	(0.1226750)
Medical offices	0.4297357	0.4425862	0.4716371	0.4673243	0.4103141
	(0.034179)	(0.052992)	(0.034259)	(0.055080)	(0.0726537)
Attorneys	0.3003019	0.3031543	0.3074149	0.3134243	0.2921774
	(0.029012)	(0.060019)	(0.029032)	(0.058330)	(0.0625478)
Log (Average	−0.0380756	−0.0513443	–	–	–
Purchase Amount)	(0.010508)	(0.013759)			
M1[a]	–	–	0.1811062	0.1980594	0.1478682
			(0.034122)	(0.044355)	(0.0456894)
M2[b]	–	–	0.4416289	0.4453912	0.3926779
			(0.029336)	(0.042118)	(0.0335346)
M3[c]	–	–	0.3266383	0.3263928	0.2724078
			(0.031110)	(0.045612)	(0.0383798)
M4[d]	–	–	0.2807466	0.2927877	0.2376187
			(0.036344)	(0.046209)	(0.0530984)
α		1.5847371		1.6192307	1.618966
		(0.0159290)		(0.017367)	(0.035522)

[a] M1 dummy variable indicates average purchase amount less than 50 dollars.
[b] M2 dummy variable indicates average purchase amount less than 100 dollars.
[c] M3 dummy variable indicates average purchase amount less than 150 dollars.
[d] M4 dummy variable indicates average purchase amount less than 200 dollars.
[e] Standard error and posterior standard deviation are in parentheses.
[f] Continuous covariates used in model specification.
[g] Dummy variable covariates used in model specification.

histories to conduct predictive tests. The sample comprises 4795 firms that were randomly selected from company records. Covariates used in the analysis include the geographic region of the firms, the type of business, and the average purchase amount. In order to fit the requirement of HB model, the average purchase amount is receded as dummy variables in increments of \$50 up to \$200 and above. In total, there were ten dummy variables available for analysis in HB model (see Table 1(a)). The average purchase frequency across all firms is equal to 4.51 per year.

Table 1(b). Poisson, NBD, and Hierarchical Bayes Parameter Estimates for Purchase Frequency (Based on the Dataset of Consumer Direct Marketing Firm)

Parameters	Poisson Model (1)	NBD Model (1)	Poisson Model (2)	NBD Model (2)	HB model
Intercept	0.6889361	−1.6679081	3.1740350	3.1718070	3.1787885
	(0.063935)[d]	(0.443480)	(0.013552)	(0.063229)	(0.0579631)
Sex	−0.0369445	−0.0310607	−0.0564288	−0.0326018	−0.0258470
	(0.102876)	(0.048619)	(0.010285)	(0.046062)	(0.0454367)
Area 1	−0.0030124	−0.0420992	−0.0139500	−0.0841114	−0.0650743
	(0.154778)	(0.072308)	(0.015485)	(0.072142)	(0.0692164)
Area 2	0.0289089	−0.0052901	−0.0105712	−0.04119727	−0.0313596
	(0.017434)	(0.086207)	(0.017424)	(0.086384)	(0.0789550)
Area 3	0.0290510	0.0436171	−0.0034828	0.0110694	0.0092028
	(0.012576)	(0.059157)	(0.012585)	(0.054018)	(0.0584529)
Area 4	−0.2127605	−0.1800427	−0.2327633	−0.2136312	−0.1687566
	(0.027654)	(0.137138)	(0.027680)	(0.136300)	(0.1096664)
Log (Average Purchase Amount)	0.2675284 (0.006907)	0.5263634 (0.048668)	–	–	–
M1[a]	–	–	−0.8596766	−0.8597417	−0.7633183
			(0.019210)	(0.071179)	(0.0619591)
M2[b]	–	–	0.0989403	0.0967286	0.1261607
			(0.013993)	(0.066687)	(0.0568964)
M3[c]	–	–	0.3855048	0.3908373	0.4012189
			(0.015336)	(0.089264)	(0.0768399)
α		0.8810592		0.9550396	0.951763
		(0.047829)		(0.043521)	(0.030209)

[a] M1 dummy variable indicates average purchase amount less than 5,000 NT dollars.
[b] M2 dummy variable indicates average purchase amount less than 10,000 NT dollars.
[c] M3 dummy variable indicates average purchase amount less than 15,000 NT dollars.
[d] Standard error and posterior standard deviation are in parentheses.

A direct marketing company, located in Taiwan and specializing in cosmetics, shampoo, toothpaste and food supplements, provided the second dataset. In contrast to the first dataset, purchases are made by individual consumers but not by other firms. The average purchase frequency is equal to 22.38 per year, allowing examination of the performance of the models with longer purchase histories. This dataset also spans a three-year period, and comprises 1725 customers randomly selected from company records. Covariates examined include the gender of the customer, geographic location and the average purchase amount. Again, the average purchase amount is recoded as dummy variables in increments of 5000 NT dollars up to 15,000 NT dollars (see Table 1(b)). In total, there were eight dummy variables available for analysis in the second dataset.

Parameter estimates for the Poisson, NBD and HB models with either continuous or dummy average purchase amount covariate are reported in Tables 1(a) and 1(b). The Poisson and NBD models were estimated by the method of maximum likelihood. Recall that for the Poisson model, the intensity parameter, λ, is linked to the covariates through the log linear format, $\lambda_i = \exp(x_i'\beta)$, while for the NBD model, $E[\lambda_i|\alpha, \theta_i] = \exp(x_i'\beta)$. In the HB model it is more natural to relate θ_i, the scale parameter of the gamma distribution, to the covariates through a log-linear model: $\theta_i = \exp(x_i'\beta)$. Therefore it is expected that the coefficient estimates for the Poisson and NBD models should agree with each other, but the intercept for the HB model will be different when α is different from one. This is true because in the gamma distribution $E[\lambda_i|\alpha, \theta_i] = \alpha\theta_i = \exp(\ln(\alpha) + x'\beta)$ implying that slope coefficient estimates should not be affected.

In practice, marketers would wish to incorporate all possible covariates, both continuous and discrete, into their prediction models. The only restriction of HB model is that all covariates must be binary coded. Thus, practitioners may be concerned about the trade-off between using a simpler model (e.g., NBD model) without converting continuous covariates, such as average purchase amount in our study, into dummy variables and using a more sophisticated model (e.g., HB model) with only dummy covariates. To investigate the possible effects of the restriction of HB model, we consider two different Poisson and NBD models. The models with continuous and dummy average purchase amount are labeled as model (1) and (2) in the tables, respectively.

Across both datasets we find close agreement in the coefficient estimates for the Poisson, NBD and HB models. The over-dispersion afforded by the NBD and HB models does not lead to tangible differences in the point estimates. Furthermore, we note that the prior distributions used in the Bayesian model (Equation (6)) do not exert much influence on the posterior estimates relative to estimates based entirely on the likelihood. We conclude that the three models provide equivalent information about the distribution of purchase frequencies and their relationship to the covariates.

The advantage of the HB model, however, is its ability to produce unique individual-level estimates of purchase frequency while incorporating covariates into the analysis. The covariates lead to individual HB estimates that differ from the conditional NBD models, particularly in small samples. To illustrate the difference in detail, Tables 2(a) and 2(b) presents individual-level estimates of purchase frequency for the Poisson, NBD, conditional NBD and HB models. Table 2(a) reports results for the firm engaged in business-to-business marketing, while Table 2(b) reports results for the firm selling consumer products. Sixteen selected customers (firms and individuals) were selected from each dataset. The selection procedure involves two steps. First, we identified four groups for each dataset based on covariates. For example, the first group in Table 2(a) consists of medical firms located in the south area of the US with average purchase amount less than \$50. Second, within each group, we selected two subjects with small purchase frequency and another two with large frequency. Customer identification numbers and associated covariates are display on the left side of the table. The observed number of purchases is reported in the center of the table, and the right side of the table contains various estimates of the customer's expected purchase frequency.

The purchase frequency estimates indicate that the HB and conditional NBD models produce estimates that are much more closely related to the observed purchase frequency in the data. Moreover, we find minor differences in the conditional NBD model when employing continuous versus dummy-variable coding of the covariates. Difference between the HB and conditional NBD estimates are greatest when the observed purchase frequency

Table 2(a). Comparison of Individual Purchase Frequency Estimates (Based on the Dataset of Business-to-Business Marketing Firm)

		Covariates				Expected Purchase Frequency[a]						
Group	Customer No.	Location	Business Type	Average Purchase Amount	Observed Purchase Frequency	Poisson Model (1)	NBD Model (1)	Conditional NBD (1)[b]	Poisson Model (2)	NBD Model (2)	Conditional NBD (2)[b]	HB
1	4772	South	Medical	M1	1	6.6873	6.8637	2.0999	5.7842	5.7852	2.0467	1.9218
	1168	South	Medical	M1	1	6.6652	6.8331	2.0981	5.7842	5.7852	2.0467	2.0331
	1938	South	Medical	M1	8	6.6733	6.8443	7.7827	5.7842	5.7852	7.5155	7.3234
	3162	South	Medical	M1	8	6.4942	6.5978	7.7284	5.7842	5.7852	7.5155	7.3983
2	2808	Midwest	Attorney	M2	1	5.8674	5.9026	2.0377	6.5729	6.6636	2.1074	2.0726
	2660	Midwest	Attorney	M2	1	5.8650	5.8994	2.0374	6.5729	6.6636	2.1074	2.2141
	2880	Midwest	Attorney	M2	10	5.8532	5.8834	9.1265	6.5729	6.6636	9.3476	9.2434
	2665	Midwest	Attorney	M2	10	5.9511	6.0165	9.1695	6.5729	6.6636	9.3476	9.3450
3	1152	Northeast	Insurance	M2	1	5.0188	5.0315	1.9656	5.6555	5.5335	2.0265	1.9338
	1392	Northeast	Insurance	M2	1	5.0632	5.0916	1.9712	5.6555	5.5335	2.0265	2.0234
	277	Northeast	Insurance	M2	19	5.0447	5.0665	15.6802	5.6555	5.5335	15.9508	15.5018
	550	Northeast	Insurance	M2	19	5.0840	5.1199	15.7192	5.6555	5.5335	15.9508	15.8712
4	4474	West	Others	M4	1	4.4660	4.4354	1.9043	4.2869	4.3091	1.9040	1.7467
	3910	West	Others	M4	1	4.4581	4.4247	1.9031	4.2869	4.3091	1.9040	1.9999
	4255	West	Others	M4	23	4.4481	4.4113	18.0871	4.2869	4.3091	17.8938	18.0019
	3300	West	Others	M4	23	4.4729	4.4446	18.1230	4.2869	4.3091	17.8938	18.1199

[a] Models labeled (1) contain continuous purchase amount in log form; models labeled (2) contain dummy purchase amount variables.
[b] Conditional NBD estimates are updated with the individual purchase information from the NBD model (i.e., Equations (3) and (4)).

Table 2(b). Comparison of Individual Purchase Frequency Estimates (Based on the Dataset of Consumer Direct Marketing Firm)

		Covariates				Expected Purchase Frequency						
Group	Customer No.	Gender	Location	Average Purchase Amount	Observed Purchase Frequency	Poisson Model (1)	NBD Model (1)	Conditional NBD (1)	Poisson Model (2)	NBD Model (2)	Conditional NBD (2)	HB
1	453	Male	North	M1	1	16.3239	11.9639	1.7520	9.4308	8.9833	1.7671	1.6926
	1723	Male	North	M1	1	18.1462	14.7333	1.7749	9.4308	8.9833	1.7671	1.7607
	468	Male	North	M1	17	17.6966	14.0238	16.8241	9.4308	8.9833	16.2296	16.2083
	270	Male	North	M1	17	18.5117	15.3228	16.9088	9.4308	8.9833	16.2296	16.2916
2	1604	Female	Central	M2	2	23.3961	22.5815	2.7729	26.1122	25.2133	2.8472	2.7505
	165	Female	Central	M2	2	22.5610	21.0231	2.7652	26.1122	25.2133	2.8472	2.9392
	982	Female	Central	M2	23	20.6236	17.6188	22.7437	26.1122	25.2133	23.0808	23.2370
	1088	Female	Central	M2	23	21.0952	18.4230	22.7909	26.1122	25.2133	23.0808	23.3580
3	1406	Male	South	M3	1	25.2412	28.8484	1.8253	33.1031	34.5070	1.9024	1.7654
	1709	Male	South	M3	1	25.2040	28.7646	1.8252	33.1031	34.5070	1.9024	1.9045
	25	Male	South	M3	68	23.5003	25.0643	66.5420	33.1031	34.5070	67.0980	66.7299
	213	Male	South	M3	68	24.8778	28.0368	66.7824	33.1031	34.5070	67.0980	67.4653
4	1393	Female	East	M1	2	15.7123	13.9371	2.7098	8.0172	8.1549	2.6452	2.6229
	1395	Female	East	M1	2	14.0243	11.1445	2.6700	8.0172	8.1549	2.6452	2.7568
	198	Female	East	M1	18	14.3461	11.6532	17.5539	8.0172	8.1549	16.9679	16.9645
	1240	Female	East	M1	18	14.8859	12.5315	17.6408	8.0172	8.1549	16.9679	17.0804

is small. For example, in Table 2(a), the rightmost columns for group 4 indicate that expected purchase frequencies may differ by as much as 10% (i.e., 1.9040/1.7467).

The difference between conditional NBD and HB models is the ability to consider the unobserved heterogeneity of individual purchasing behavior. The updating schema for conditional NBD will produce identical estimates for those who have the same purchase frequency. In contrast, the HB estimator takes both observed and unobserved heterogeneity into account, and can result in different estimates for customers with identical purchase frequency.

Summary measures of the in-sample and predictive fit of the various models were calculated by re-estimating the model on a subset of the original data and by using the remainder as a holdout. Both the business-to-business and consumer direct marketing data sets span a total of three years. The first half of each customer's purchase history (spanning 1.5 years) was used to assess the in-sample fit of the model, and the second half was used to assess out-of-sample fit.

Table 3 reports the in-sample fit of various models used to estimate expected purchase frequency. Two measures of fit are reported, the root mean squared error (RMSE) and mean absolute deviation (MAD) between the fitted values and the observed frequency. Also reported is the correlation between the expected and observed frequencies. This latter statistic can be used to assess the accuracy of an implied ranking of the firms based on expected purchase frequency. The results reported in Table 3 indicate that the HB estimates have highest predictive accuracy for the business-to-business dataset in which the customers have small purchase frequency on average. However, in the case of consumer direct marketing dataset where the observed purchase frequency is large for most households, the relative performance of these two models is somewhat mixed. Although the conditional NBD model with *continuous* covariates is slightly better than HB model on all criteria, HB model still outperforms the conditional NBD model with *discrete* covariates in terms of RMSE and correlation criteria.

Table 3. In-Sample Fit (Percentage Improvement of HB Model in Parentheses)

	Poisson Model (1)	NBD Model (1)	Conditional NBD (1)	Poisson Model (2)	NBD Model (2)	Conditional NBD (2)	*HB Model*
				Based on the Dataset of Business-to-Business Marketing Firm			
RMSE	2.7971	2.8536	1.2282	2.7902	2.7930	1.2464	*0.8681*
	(69%)	(70%)	(29%)	(69%)	(69%)	(30%)	
MAD	1.5749	1.6195	0.7467	1.6354	1.6346	0.7532	*0.5213*
	(67%)	(68%)	(30%)	(68%)	(68%)	(31%)	
Correlation	0.3855	0.3671	0.9552	0.3843	0.3821	0.9589	*0.9833*
				Based on the Dataset of Consumer Direct Marketing Firm			
RMSE	11.6487	14.8722	0.8182	10.6769	10.7501	0.9377	*0.8439*
	(93%)	(94%)	(−3%)	(92%)	(92%)	(10%)	
MAD	6.4543	7.2683	0.4361	6.4888	6.5353	0.4263	*0.4389*
	(93%)	(94%)	(−0.6%)	(93%)	(93%)	(−2.9%)	
Correlation	0.5545	0.4228	0.9991	0.6411	0.6348	0.9981	*0.9987*

Table 4. Predictive Performance Based on Different Amount of Information (Percentage Improvement of HB Model in Parentheses)

	Poisson Model (1)	NBD Model (1)	Conditional NBD (1)	Poisson Model (2)	NBD Model (2)	Conditional NBD (2)	*HB Model*
			Based on the Dataset of Business-to-Business Marketing Firm				
RMSE							
1. Based on 6 months buying records	4.1641 (10%)	4.2309 (11%)	4.1218 (9%)	3.9183 (4%)	3.9188 (4%)	3.9853 (5.6%)	*3.7640*
2. Based on 12 months buying records	3.8213 (9%)	3.8959 (11%)	3.6176 (4%)	3.7152 (6%)	3.7136 (6%)	3.5784 (2.6%)	*3.4848*
3. Based on 18 months buying records	3.6788 (14%)	3.7418 (16%)	3.2288 (2%)	3.6354 (13%)	3.6340 (13%)	3.2272 (2.3%)	*3.1515*
MAD							
1. Based on 6 months buying records	2.3817 (21%)	2.4392 (23%)	2.5258 (26%)	2.0176 (7%)	2.0181 (7%)	2.3283 (19.2%)	*1.8812*
2. Based on 12 months buying records	2.0878 (21%)	2.1918 (25%)	1.9872 (18%)	1.8297 (10%)	1.8261 (10%)	1.8309 (10.5%)	*1.6390*
3. Based on 18 months buying records	1.9329 (25%)	2.0395 (29%)	1.6599 (13%)	1.7214 (16%)	1.7159 (16%)	1.5400 (6.0%)	*1.4470*
			Based on the Dataset of Consumer Direct Marketing Firm				
RMSE							
1. Based on 6 months buying records	17.9064 (6%)	20.3277 (17%)	16.8651 (0.55%)	17.1429 (2%)	17.3474 (3%)	16.9921 (1.3%)	*16.7722*
2. Based on 12 months buying records	17.1005 (9%)	19.3509 (20%)	15.5556 (0.03%)	16.4154 (5%)	16.5565 (6%)	15.6148 (0.4%)	*15.5510*
3. Based on 18 months buying records	16.4660 (15%)	18.5429 (25%)	13.9476 (0.16%)	15.6741 (11%)	15.6934 (11%)	13.9322 (0.1%)	*13.9247*
MAD							
1 . Based on 6 months buying records	12.2572 (12%)	13.0978 (17%)	11.3539 (4.56%)	11.4773 (6%)	11.6489 (7%)	11.1662 (3.0%)	*10.8358*
2. Based on 12 months buying records	11.8460 (14%)	12.7708 (20%)	10.3929 (2.21%)	11.0827 (8%)	11.2290 (9%)	10.2701 (1.0%)	*10.1634*
3. Based on 18 months buying records	11.2458 (20%)	12.1809 (26%)	9.1781 (1.60%)	10.4779 (14%)	10.5073 (14%)	9.0690 (0.4%)	*9.0310*

Table 4 reports out-of-sample fits of the models under various lengths of the customer's purchase history. We use either the first 6, 12 or 18 months of records in the first half of the customer purchase history to produce estimates of expected purchase frequency, and then compare these estimate to observed frequencies in the second half of the data. We expect that with very limited customer purchase histories, covariates will provide a useful basis for generating estimates of purchase frequency. As the length of the purchase history increases, individual-level estimates that rely on the customer's observed purchase rate should yield more accurate estimates.

The predictive performance of the HB model is better than all other models, with the increase in predictive accuracy greatest with small sample sizes. In comparison to the conditional NBD model, which simply adjusts expected purchase frequency with the observed frequency, the HB estimates are more accurate when there exists limited individual-level information. For the business-to-business dataset, which has relatively infrequent customer interactions, the HB estimates are 9 and 5.6% more accurate than the conditional NBD model (1) and (2), respectively, when forecasts are based on 6 months of buying records. As the length of the purchase history increases, the conditional NBD and HB estimates converge, and offer substantial improvement relative to the aggregate Poisson and NBD models.

The hierarchical Bayes model offers a compromise between standard models (Poisson and NBD) that pool purchases across customers and relate them to covariates such as demographics, and the conditional NBD model that adjusts expected purchase frequency for each customer in isolation of the others by simply using a linear transformation formula. The HB estimator retains the flexibility of the conditional NBD model and produces individual-level estimates without imposing undue restrictions on the variability of these estimates as encountered with the standard models. The HB estimator is a compromise between the other estimators, offering the same in-sample fit as the conditional NBD estimates and equivalent predictive fits as the standard models when there is no data available for the new customers. In addition, the HB estimates show a tendency of convergence as the sample size at the individual level increases (Huber and Train, 2001). For the consumer direct marketing dataset with high purchase frequency, there is no major difference between the conditional NBD and the HB models. It is when the sample size is low as in the business-to-business dataset that the HB model does better.

3. Concluding Remarks

This paper introduces a hierarchical Bayes model of purchase frequency with observed and unobserved components of heterogeneity. Unobserved heterogeneity is captured through a gamma distribution for purchase rates, while observed heterogeneity is introduced by allowing the location-scale parameter of the distribution to depend on covariates. We demonstrate the advantage of using hierarchical Bayes approach relative to the other approaches, and provide algorithms for estimating the model with the Gibbs sampler. 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.

We examine HB estimates of expected purchase frequency in two datasets. Based on these results we expect the proposed methodology to yield improved estimates of expected frequency relative to conventional approaches, particularly when there exists limited information about any particular customer. This is because the hierarchical Bayes estimator augments the limited information contained in the customer purchase history with information from the random-effects distribution and information contained in available covariates.

There are a number of avenues for future work in this area. 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 that a constant hazard rate which may be overly restrictive in many circumstances. However, violations of this assumption would require longer purchase histories than that present in the business-to-business direct marketing problem analyzed in this paper (see Allenby et al. (1999) for a model of purchase timing spanning six years). Second, we allow the location-scale parameter of the gamma distribution to be related to indicator (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-scale parameter to be directly related to continuous covariates such as age or income. Finally, we note that customer scoring systems in direct marketing also use variables such as the monetary amount of the purchase to estimate the total worth of their customers. Extending the model to include multivariate response variables (e.g., inter-purchase time, monetary amount) would therefore be a fruitful avenue for future research.

Appendix: The Gibbs Sampler for Hierarchical Bayes Model of Purchase Rate

Denote y_i as the number of purchases made by the ith customer in a time period of length T. Assume y_i is distributed Poisson with gamma heterogeneity:

- likelihood:

$$\ell(Y_i = y_i|\lambda_i, T) \propto \lambda_i^{y_i} \mathrm{e}^{-\lambda_i T}, \tag{A1}$$

- mixing distribution:

$$\pi(\lambda_i|\alpha, \theta_i) = \frac{\lambda_i^{\alpha-1}}{\Gamma(\alpha)\theta_i^{\alpha}} \mathrm{e}^{-(\lambda_i/\theta_i)}, \quad \text{where } \theta_i = \exp(x_i'\beta). \tag{A2}$$

The Gibbs sampler recursively generates draws from the full conditional distributions of the model. Below we derive the conditional distributions for the Poisson–gamma model.

Estimation Algorithm

1. *Generate* $\lambda_i,\ i = 1, \ldots, N$ (one unit at a time)

$$\pi[\lambda_i|y_i, T, \alpha, x_i, \beta] \propto \lambda_i^{y_i} \mathrm{e}^{-\lambda_i T} \lambda_i^{\alpha-1} \mathrm{e}^{-\lambda_i[\exp(x_i'\beta)]^{-1}} \propto \lambda_i^{y_i+\alpha-1} \mathrm{e}^{-\lambda_i\{T+[\exp(x_i'\beta)]^{-1}\}}.$$

   Consequently, λ_i is generated from a gamma distribution, $\mathrm{G}(A, B)$, where

$$A = Y_i + \alpha \quad \text{and} \quad B = \left(T + \left[\exp(x_i'\beta)\right]^{-1}\right)^{-1}, \quad \text{for } i = 1, \ldots, N.$$

2. *Generate* $\beta_k,\ k = 1, \ldots, K$ (one coefficient at a time)
   The posterior distribution of β_k, given the other parameters, is

$$\pi\left(\beta_k|\{\lambda_i\}, \alpha, x_{ik}, T\right) \propto \prod_{i=1}^{N} \ell(\lambda_i|\alpha, \beta_k, x_{ik}, T)\pi(\beta_k).$$

In order to generate β_k, we first re-parameterize β_k by letting $e^{\beta_k} = \varphi_k$ and then derive the posterior distribution of φ_k. We also assume that all the covariates x_k, where $k \in (1, \ldots, K)$, are 0–1 dummy variables. Since

$$e^{x_i'\beta} = e^{\sum_{j=1}^{K} x_{ij}\beta_j} = \prod_{j=1}^{K} \left(e^{\beta_j}\right)^{x_{ij}} = \prod_{j=1}^{K} \varphi_j^{x_{ij}}.$$

Equation (A2) can be rewritten as

$$\pi\left(\lambda_i|\alpha, \exp(x_i'\beta)\right) = \frac{\lambda_i^{\alpha-1}}{\Gamma(\alpha)\left[\prod_{j=1}^{K} \varphi_j^{x_{ij}}\right]^{\alpha}} \cdot e^{-\lambda_i/\prod_{j=1}^{K} \varphi_j^{x_{ij}}}.$$

Define $\sum_{i=1}^{N} x_{ik} = N_k$, $S_{(k)} = \{x_{ik}|x_{ik} = 1,\ i = 1, \ldots, N\}$, and $D_i^{(k)} = \prod_{j=1, j\neq k}^{K} \varphi_j^{x_{ij}}$. The likelihood of β_k in terms of φ_k is

$$\begin{aligned}
\ell(\varphi_k|x_i, \lambda_i, T) &\propto \left(\prod_{i=1}^{N} \varphi_1^{-\alpha x_{i1}}\right)\left(\prod_{i=1}^{N} \varphi_2^{-\alpha x_{i2}}\right)\cdots\left(\prod_{i=1}^{N} \varphi_k^{-\alpha x_{ik}}\right)\cdots\left(\prod_{i=1}^{N} \varphi_K^{-\alpha x_{iK}}\right) \\
&\quad \times \exp\left[-\sum_{i=1}^{N}\left(\varphi_1^{-x_{i1}} \cdot \varphi_2^{-x_{i2}} \cdots \varphi_k^{-x_{ik}} \cdots \varphi_K^{-x_{iK}} \cdot \lambda_i\right)\right] \\
&\propto \varphi_k^{-\alpha \sum_{i=1}^{N} x_{ik}} \cdot \exp\left[-\varphi_k^{-1} \sum_{i\in S_{(k)}} \lambda_i \left(D_i^{(k)}\right)^{-1} + \sum_{i\notin S_{(k)}} \lambda_i \left(D_i^{(k)}\right)^{-1}\right] \\
&\propto \varphi_k^{-\alpha N_k} \cdot \exp\left[-\varphi_k^{-1} \sum_{i\in S_{(k)}} \lambda_i \left(D_i^{(k)}\right)^{-1}\right].
\end{aligned}$$

If the prior distribution of φ_k is Inverse Gamma distribution, $\mathrm{IG}(a_{k0}, b_{k0})$, the conditional posterior of φ_k is proportional to

$$\begin{aligned}
&\pi\left(\varphi_k|\lambda_i, x_{ik}, D_i^{(k)}, a_{k0}, b_{k0}, T\right) \\
&\quad \propto \varphi_k^{-(\alpha N_k + a_{k0}+1)} \exp\left[-\varphi_k^{-1}\left(b_{k0}^{-1} + \sum_{i\in S_{(k)}} \lambda_i \left(D_i^{(k)}\right)^{-1}\right)\right],
\end{aligned}$$

then generate φ_k from

$$\text{Inverse Gamma}\left(\alpha N_k + a_{k0}, \left(b_{k0}^{-1} + \sum_{i\in S_{(k)}} \lambda_i \left(D_i^{(k)}\right)^{-1}\right)^{-1}\right).$$

3. *Generate α*

The posterior distribution of α, given the other parameters, is

$$\pi\left(\alpha|\{\lambda_i\}, \beta_k, x_{ik}, T\right) \propto \prod_{i=1}^{N} \ell(\lambda_i|\alpha, \theta_i)\pi(\alpha).$$

Based on the Equation (A2), the likelihood of α is:

$$\ell(\alpha|\lambda_i, \theta_i, T) \propto \prod_{i=1}^{N} \frac{\lambda_i^{\alpha-1}}{\Gamma(\alpha)\theta_i^{\alpha}} \mathrm{e}^{-\lambda_i/\theta_i} = \prod_{i=1}^{N} \frac{(\lambda_i/\theta_i)^{\alpha}}{\Gamma(\alpha)\lambda_i} \mathrm{e}^{-\lambda_i/\theta_i}.$$

If the prior distribution of α is Uniform (M), then the posterior distribution of α, given λ_i and θ_i, is proportional to $\prod_{i=1}^{N}$ *continuous Poisson* (λ_i/θ_i), which is the product of N continuous Poisson distributions.

References

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

Blattberg, Robert, Gary Getz, and Jacquelyn Thomas. (2001). *Customer Equity: Building and Managing Relationships as Valuable Assets*. Boston: Harvard Business School Press.

Brockett, Patrick, Linda Golden, and Harry Panjer. (1996). "Flexible Purchase Frequency Modeling," *Journal of Marketing Research*, 33(1), 94–107.

Bucklin, Randolph, Sunil Gupta, and S. Siddarth. (1998). "Determining Segmentation in Sales Response Across Consumer Purchase Behaviors," *Journal of Marketing Research*, 35(2), 189–197.

Cameron, Colin, and Pravin Trivedi. (1986). "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests," *Journal of Applied Econometrics*, 1(1), 29–54.

David Shepard Associates, Inc. (1995). *The New Direct Marketing: How to Implement a Profit-Driven Database Marketing Strategy*. New York: Richard D. Irwin, Inc.

Ehrenberg, A. S. C. (1959). "The Pattern of Consumer Purchases," *Applied Statistics*, 8(1), 26–41.

Ehrenberg, A. S. C. (1988). *Repeat Buying: Theory and Applications*. New York: Oxford University Press.

Gelfand, Alan, and Adrian Smith. (1990). "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.

Grogger, J. T., and Richard T. Carson. (1991). "Models For Truncated Counts," *Journal of Applied Econometrics*, 6(3), 225–238.

Hausman, Jerry, Bronwyn Hall, and Zvi. Griliches. (1984). "Econometric Models for Count Data With an Application to the Patents–R&D Relationship," *Econometrica*, 52(4), 909–938.

Huber, Joel, and Kenneth Train. (2001). "On the Similarity of Classical and Bayesian Estimates of Individual Mean Partworths," *Marketing Letters*, 12(3), 259–269.

Massy, William, David Montgomery, and Donald Morrison. (1970). *Stochastic Models of Buying Behavior*. Cambridge: The MIT Press.

Michener, Ron, and Carla Tighe. (1992). "A Poisson Regression Model of Highway Fatalities Externalities," *The American Economic Review*, 82(2), 452–456.

Morrison, Donald, and David Schmittlein. (1988). "Generalizing the NBD Model for Customer Purchases: What Are the Implications and Is it Worth the Effort?," *Journal of Business and Economic Statistics*, 6(2), 145–159.

Shaw, Daigee. (1988). "On-Site Samples' Regression: Problems of Non-Negative Integers, Truncation, and Endogenous Stratification," *Journal of Econometrics*, 37(2), 211–223.

Stone, Bob. (1994). *Successful Direct Marketing*. Lincolnwood, IL: NTC Business Books.