Journal of Marketing Research
Article Postprint
Volume XLV
© 2008, American Marketing Association
Cannot be reprinted without the express
permission of the American Marketing Association.

The Importance of Modeling Temporal Dependence of Timing and Quantity in Direct Marketing

Lichung Jen
Associate Professor
Department of International Business
National Taiwan University
1, Sec. 4, Roosevelt Rd.
Taipei, Taiwan
lichung@ntu.edu.tw

Chien-Heng Chou
Associate Professor
Department of International Trade
Chinese Culture University
55, Hwa Kang Rd.
Hwa Kang, Yang Ming Shan
Taipei, Taiwan
chouchdr@faculty.pccu.edu.tw

Greg M. Allenby
Professor
Ohio State University
2100 Neil Avenue
Columbus, OH 43210
allenby.1@osu.edu

April, 2008

Author Note

The authors thank anonymous reviewers for their constructive comments.

Abstract

The analysis of customer value in direct marketing typically combines customer timing and quantity data into a single statistic used to compute lifetime values, rank-order of customers for differential action, and identify prospects for cross-selling. Current models, however, assume that purchase timing and quantity decisions are independently realized (i.e., uncorrelated) over time given individual-level parameters. In this paper, the authors show that customer value calculations can be severely biased in these models when timing and quantity are dependently related. The authors propose alternative models that are shown to lead to substantial gains in profitability in two direct marketing datasets. The results indicate that the commonly held assumption of independence leads to an over-valuation of customer value.

Keywords

CRM, customer scoring, lifetime value, covariance heterogeneity, Bayesian analysis

The Importance of Modeling Temporal Dependence of Timing and Quantity in Direct Marketing

The emergence of customer relationship management (CRM) with its focus on individual customers has led to the development of disaggregate models and analyses of heterogeneous behavior. In the CRM literature, customers who frequently interact with a firm are expected to generate greater income streams and have longer expected lives than those who interact infrequently (Blattberg, Getz, and Thomas 2001). Similarly, heavy buyers are expected to generate greater profits and exhibit greater loyalty than light buyers (Jacoby and Kyner 1973; Reichheld, Markey, and Hopton 2000). As a result, the value of a customer's future income stream is generally thought to be related to the variability of the timing and quantities of their orders.

Researchers in marketing have been successful in developing models of purchase quantity (Schmittlein and Peterson 1994), inter-purchase times (Jen and Wong 1998; Jen, Chou and Allenby 2003), and models of purchase quantities and timing (Boatwright, Borel and Kadane 2003), but have not been successful in developing scoring models that integrate these measures while allowing for their heterogeneous covariation. If a customer compensates for late orders by ordering larger quantities, the income stream from that individual would remain constant and their timing and quantity statistics would exhibit positive covariation. In contrast, customers who do not exhibit compensating behavior will exhibit smaller covariation between their timing and quantity of orders, but may have income streams with greater variability.

The presence of non-zero covariation between purchase quantity and timing is at odds with the majority of model computing customer lifetime value. Purchase quantity and timing are assumed to be temporally independent in the models of Schmittlein and Peterson (1994),

Colombo and Jiang (1999), Reinartz and Kumar (2000, 2003), Rust, Lemon and Zeithaml (2004), and Fader, Hardie and Lee (2005), among others. If these quantities are not independent, predictions of customer value will be biased. We note that the presence of temporally dependent variables cannot be "corrected" with correlated random-effects, or some other cross-sectional model (e.g., Borel, Singh and Jain 2008). Any model correction, or re-specification, must take place at the individual-level.

To illustrate, let customer value be denoted as y/time , where " y " denotes purchase quantity. Customer value can then be expressed as:

$$E\left[\frac{y}{\text{time}}\right] = E\left[\exp(\ln y - \ln \text{time})\right] \quad (1)$$

A series expansion of the exponential function is

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots \quad (2)$$

Applying the series expansion to equation (1) yields:

$$\begin{aligned} E\left[\exp(\ln y - \ln \text{time})\right] &= E\left[1 + (\ln y - \ln \text{time}) + \frac{(\ln y - \ln \text{time})^2}{2!} + \dots\right] \\ &= 1 + E(\ln y - \ln \text{time}) + \frac{\text{Var}(\ln y - \ln \text{time}) + [E(\ln y - \ln \text{time})]^2}{2!} \end{aligned} \quad (3)$$

Expected customer value is therefore an increasing function of the variance of log quantity minus log time. Since $\text{Var}(A-B) = \text{Var}(A) + \text{Var}(B) - 2\text{Cov}(A,B)$, an incorrect assumption of independence *over-values* customers with compensating purchase behavior (i.e., behavior with positive covariation).

In this paper we develop a hierarchical model of purchase timing and quantity that allows for the covariation of these quantities over time and among respondents. We propose a parameterization of the bivariate model that overcomes restrictive conditions encountered with modeling heterogeneous covariances, and show that this parameterization leads to improved prediction and inference in two direct marketing datasets. We begin in the next section by developing the model and comparing it to existing models with heterogeneous covariances (Ansari, Essegai, and Kohli 2000). Two empirical applications are provided, and alternative versions of the proposed model are compared to existing models in terms of fit and predictive accuracy. We then investigate the ability of proposed model to integrate purchase timing and quantity into a score that more accurately reflects a customer's expected income stream. We demonstrate large financial gains from using the proposed model to identify high-value customers. Concluding remarks are then offered in the last section.

MODEL DEVELOPMENT

We develop a model for customer scoring that integrates interpurchase time (t) and purchase amount (y) within the context of a bivariate log-normal distribution. An advantage of our specification is that it facilitates calculations of customer value expressed as $E[y/t]$, the expected revenue stream. In previous studies that treat these two variables independently, purchase monetary amount has been assumed to follow either a normal (e.g., Schmittlein and

Peterson 1994) or a gamma distribution (e.g., Fader, Hardie, and Lee 2005), while interpurchase time has always been assumed to be a gamma variable (e.g., Allenby, Leone, and Jen 1999; Venkatesan and Kumar 2004). However, due to the fact that both variables have a nonnegative domain, it is logical to assume these variables to be log-normally distributed. Moreover, since the covariance of these two variables is the focal point of this research, the joint distribution of these two random variables can be specified as a bivariate log-normal distribution.

Let y_{ij} be the logarithm of the j^{th} purchase amount of the i^{th} customer, and t_{ij} be the logarithm of the j^{th} interpurchase time, where $i=1, 2, \dots, N$ and $j=1, 2, \dots, n_i$. Then, we have

$$\begin{bmatrix} y_{ij} \\ t_{ij} \end{bmatrix} \sim \text{Normal}(\mu_i, \Sigma_i) \quad (3)$$

and the distribution of log value is easily obtained as the $\log(y/\text{time}) = y_{ij} - t_{ij} \sim \text{Normal}(\mu_{y,i} - \mu_{t,i}, \sigma_{yy,i} + \sigma_{tt,i} - 2\sigma_{yt,i})$ using properties of the normal distribution. We note that since the model allows for heterogeneity in the mean vector and covariance matrix of the bivariate normal distribution, it can reflect heterogeneous behavior in terms of levels (μ_i) and variability (Σ_i), including covariance $\sigma_{yt,i}$.

When sufficient data exists at the individual level, model parameters μ_i and Σ_i can be estimated using maximum likelihood techniques. However, it is typically the case that, at the individual-level, there exists limited amounts of data (Hughes 2005; Miglautsch 2002). When this occurs, hierarchical Bayes (HB) models have been proposed to improve estimates of individual-level parameters (see Rossi and Allenby 2003 for an overview).

A standard hierarchical Bayes model for the normal distribution is to assume that the mean vector, μ_i , is distributed normal across customers, and to employ an inverted Wishart distribution for Σ_i (Ansari, Essegiaier, and Kohli 2000; Ying, Feinberg, and Wedel 2006). Although the model specification and associated estimation algorithms of the standard HB model are well established, a drawback of employing this model arises when using the inverted Wishart distribution to model heterogeneity of the covariance matrix across respondents. The inverted Wishart distribution contains a single degree-of-freedom parameter that applies to all elements of the covariance matrix that we show to be restrictive in our analysis reported below.

The single degree-of-freedom parameter would not raise any problems in large samples because the likelihood dominates the prior. But, in samples of the small size (e.g., less than 5 observations) found in many marketing and RFM applications, the inverted Wishart distribution plays an important role in augmenting the limited information that exists at the individual customer level, and greater flexibility is required in the covariance heterogeneity distribution.

Conditional Normal HB Model

We propose a re-parameterization of the bivariate normal distribution that allows use of a more flexible distribution of covariance heterogeneity. Researchers have found that the parameterization of the likelihood can affect inferences in Bayesian models because of the presence of the prior distribution, and, in our case, the distribution of heterogeneity. The heterogeneity distribution governs the sharing of information across units of analysis, and restrictive properties of these distributions can influence the estimated posterior distribution for individual customers. As discussed by Sonnier, Ainslie and Otter (2007) and Allenby et.al. (2005) such considerations are not present in maximum likelihood estimation where re-

parameterization of the likelihood leads to equivalent point estimates of parameters – i.e., the value of parameters that maximize the likelihood function can be exactly predicted under any re-parameterization of the likelihood (Zehna 1966)

We propose a model where the bivariate normal distribution is built-up from a marginal and conditional distribution. We name our model the "conditional normal HB model":

$$f(t_{ij}) \sim N(\nu_i, \varsigma_i^2) \quad (5)$$

$$f(y_{ij} | t_{ij}) \sim N(\mu_i + t_{ij}\eta_i, \sigma_i^2) \quad (6)$$

where "i" indexes the respondent, ν and ς^2 are parameters for the marginal distribution of the (log) purchase times, and μ , η , and σ^2 parameterize the conditional distribution of (log) purchase amount given (log) time.

This approach to building the joint model is similar to that used by Boatwright, Borle and Kadane (2003) who assume a discrete distribution (i.e., the CM-Poisson distribution) for timing and a continuous distribution for purchase quantity. The use of such a discrete/continuous model, however, creates problems in deriving inference for summary score measures such as the expected revenue per time because it does not have a convenient parameterization. An advantage of using a normal distribution is that the distribution of log value ($y_{ij}-t_{ij}$) is a simple function of model parameters as discussed above.

Our model uses a separate inverted gamma distribution for the prior on each diagonal element (i.e. σ_i^2 and ς_i^2). This results in a model specification with greater flexibility, allowing elements of the covariance matrix to vary across the population independently. Below, we show

that our proposed model is more flexible in reflecting heterogeneous patterns of purchase timing and purchase quantity.

From Equations (5) and (6), the joint distribution of y_{ij} and t_{ij} is equal to:

$$f(y_{ij}, t_{ij}) \sim N \left(\begin{bmatrix} \mu_i + \nu_i \eta_i \\ \nu_i \end{bmatrix}, \begin{bmatrix} \eta_i^2 \varsigma_i^2 + \sigma_i^2 & \eta_i \varsigma_i^2 \\ \eta_i \varsigma_i^2 & \varsigma_i^2 \end{bmatrix} \right) \quad (7)$$

and the corresponding correlation of y_{ij} and t_{ij} is

$$\rho_i = \frac{\eta_i \varsigma_i}{\sqrt{\eta_i^2 \varsigma_i^2 + \sigma_i^2}} \quad (8)$$

which is naturally constrained to be in the interval [-1,1]. While the proposed model has the same number of parameters ($\mu_i, \nu_i, \eta_i, \sigma_i, \varsigma_i$) as the standard model ($\mu_t, \mu_y, \sigma_{tt}, \sigma_{ty}, \sigma_{yy}$), it allows for a more flexible pattern of shrinkage because the parameters σ_i and ς_i are assumed to be independently distributed inverse gamma, and η_i , the parameter leading to dependence between purchase timing and quantity, is assumed to be normally distributed.

Similar to the standard HB parameterization for the bivariate normal distribution, it is possible to incorporate covariates, x , into the model structure by letting $\mu_i \sim \text{Normal}(x_i' \beta, \zeta^2)$ and $\nu_i \sim \text{Normal}(x_i' \gamma, \xi^2)$. Details of distributional assumptions and Bayesian estimation algorithms are provided in the web appendix.

Simulation Study

We investigate the performance of our proposed parameterization relative to a standard normal model using two simulated datasets. The data for the first study were simulated by generating draws of μ_i and Σ_i representing hypothetical customers where the diagonal elements of Σ_i are approximately equal.

$$\begin{bmatrix} \mu_y \\ \mu_t \end{bmatrix}_i \sim \text{Normal}\left(\begin{bmatrix} 5 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\right) \quad \Sigma_i \sim \text{IW}\left(5, \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}^{-1}\right) \quad (9)$$

The number of observations per customer, n_i , was assumed to follow a Poisson ($\lambda=6$) distribution, corresponding to the distribution of purchases in our empirical application reported below. A second dataset was generated with unequal elements of the covariance matrix:

$$\begin{bmatrix} \mu_y \\ \mu_t \end{bmatrix}_i \sim \text{Normal}\left(\begin{bmatrix} 5 \\ 2 \end{bmatrix}, \begin{bmatrix} 10 & 4 \\ 4 & 2 \end{bmatrix}\right) \quad \Sigma_i \sim \text{IW}\left(5, \begin{bmatrix} 10 & 4 \\ 4 & 2 \end{bmatrix}^{-1}\right) \quad (10)$$

The simulated data were then used to compare results from a standard bivariate normal HB model with covariance modeled with an Inverted Wishart distribution, and our conditional HB model. Results are provided in Table 1 and Figure 1. The top portion of Table 1 reports on the recovery of model hyper-parameters, and the bottom portion of the table reports summary statistics for the recovery of individual-level parameters. Both models are capable of recovering the hyper-parameters because of the large number of respondents in the simulation study. However, the proposed conditional model is better at recovering individual-level parameters. This is particularly true in the second dataset, where the diagonal elements differ. The Inverted Wishart distribution does not perform well because one degree-of-freedom parameter governs behavior of all elements of the matrix. The conditional model more flexibly recovers elements of the covariance matrix in small samples.

[Table 1]

Figure 1 illustrates the role of sample size in influencing the estimation performance of the standard bivariate model and the proposed conditional model for the second dataset. Similar results are also found with the first dataset. Plotted is the Bayesian measure of model fit, the average log marginal density, for simulated respondents with specific numbers of observations (n_i). For small samples, the standard model performs much worse than the conditional model because the Inverted Wishart prior has greater influence on the posterior distribution. As the purchase frequency increases, the two models converge in fit because the prior is less influential.

[Figure 1]

We compare variations of our conditional model to two datasets in the next section to investigate the presence of influence of non-zero covariance between purchase quantity and purchase timing, and the presence of covariance heterogeneity across customers. The parameter η_i in Equation (6) reflects the amount of covariation. If η_i is set to zero, then inter-purchase time and purchase monetary amount are assumed to be independently distributed. If the model only allows σ^2 instead of σ_i^2 (i.e. $\sigma_i^2 = \sigma^2$), then the model assumes homogeneity of variability of inter-purchase time and purchase monetary amount. Thus, four competing and nested models can be obtained: a dependent-heterogeneous model (DHT) (the model we proposed in this research), a dependent-homogeneous model (DHO), an independent-heterogeneous model (IDHT), and an independent-homogeneous model (IDHO). Homogeneity refers here to homogeneity of the covariance matrix, not homogeneity of the mean vector of the log-normal distribution.

EMPIRICAL APPLICATION

Data

Two datasets are used to evaluate the performance of the proposed models. The first dataset is from a company engaged in business-to-business marketing, selling office supply products in the United States. A sample of 1935 firms, which had at least five purchase incidents within a period of three years, was randomly selected from company records. These firms were located in four geographic areas (i.e. Northeast, Midwest, South, and West) and were classified into four types of business (i.e., Insurance-related Firms, Medical Offices, Attorneys, and Others). The four geographic regions were represented by four dummy variables, and the four types of business were represented by three dummy variables with Others as base. In total, there were seven covariates used in the following regression analysis without intercept term. The average purchase amount across all firms is equal to US\$98, with a standard deviation of 72.18, and the average inter-purchase time is 135 days, with a standard deviation of 70.47.

The second dataset comprises 1344 consumers randomly drawn from a direct marketing company in Taiwan selling health and beauty goods (e.g. cosmetics, skin care products, etc.). Similar to the selection criterion used in the first dataset, those customers who also made at least five purchases within a three-year period were included. However, in contrast to the first dataset, purchases are made by individual consumers, not firms. The average purchase amount is NT\$10,342 (equivalent to about US\$ 295), with a standard deviation of NT\$8,551. The average interpurchase time is 43 days (standard deviation equal to 45.89), which is much shorter than that of the first dataset. There were six covariates used in the analysis representing the gender and geographic location of the customers (i.e., North, West, South, East and Central). Two dummy variables were used to represent gender and four dummy variables were used to represent geographic areas with Central as base.

Estimation

Both datasets were partitioned into calibration and predictive parts, each representing roughly half the data (i.e., one and a half years), with each customer having at least two observations in each part. The first part was used to estimate model parameters, and the second half was used to evaluate the predictive performance of the models. The models were estimated using 1000 MCMC iterations, from which the first 500 iterations were discarded and the last 500 iterations were used to form estimates of the posterior distribution of model parameters. Time series plots of the draws indicate the convergence of the chain from multiple initial values. Figure 2 displays a time series plot of the parameters related to dependence (η_i) and heterogeneous covariances (σ_i^2) for a randomly selected customer. The plots show that the chain converges quickly, within about 200 iterations.

[Figure 2]

Parameter estimates for the IDHO, IDHT, DHO, DHT, and the standard bivariate normal model using the inverted Wishart distribution are reported in Tables 2 and 3. We also estimated an OLS regression model as an additional benchmark model. The top portion of the table reports parameter estimates for the marginal distribution of purchase timing (Equation 5), and estimates for the conditional distribution of purchase quantity (Equation 6) are reported in the center of the table. Covariance parameters for the bivariate normal HB model are reported at the bottom of the table. Across both datasets we find general agreement in the coefficient estimates for all of the five models. Moreover, we find that many of the covariates have posterior mass away from zero.

Of critical interest are the posterior estimates associated with the parameter η_i that provides evidence of covariation in purchase timing and amount. We find evidence of positive covariation in both datasets, with stronger evidence in the B-B data estimates reported in Table 2. Estimates of the mean and variation of the η_i parameter indicate that most customers exhibit positive covariation in purchase timing and quantity, while for the B-C customers where are more customers that also exhibit negative covariation. We also find that the β coefficient estimates tend to be smaller in the B-B data when this covariation is modeled. In contrast, evidence from the standard normal bivariate model shows weak negative covariation.

[Tables 2 and 3]

Figure 3 displays the posterior distribution of the correlation coefficient, ρ_i , between interpurchase time and purchase monetary amount among the customers in each dataset. The distribution is obtained by evaluating Equation (8) on each iteration of the chain for each customer, and then combining these draws to obtain the distribution. The mean of the ρ_i for business-to-business marketing firm is positive, indicating that customers tend to adjust their purchase monetary amount upwards when interpurchase time are longer. In contrast, the distribution of the correlation coefficient for the business-to-consumer data are concentrated near zero, indicating that the purchase patterns of most customers are not regular.

[Figure 3]

Model Fit Statistics

Predictive fits of the models are reported in Table 4. Two measures of fit, root mean squared error (RMSE) and mean absolute deviation (MAD) between the fitted and observed values of inter-purchase time (t), purchase monetary amount (y) and customer value (y/t) are reported. Also reported is the percentage improvement of our proposed dependent

heterogeneous covariance (DHT) model to the other five competing models. Across both datasets, the hierarchical Bayes models are seen to outperform the regression model. The percentage improvement in RMSE by DHT model is from 5.25% to 6.51% for business-to-business dataset, and from 6.95% to 16.13% for business-to-consumer dataset. Similar results are also shown by the measurement of MAD. This indicates that capturing heterogeneity in mean level leads to better estimates. In addition, the proposed models are seen to out-predict the bivariate normal model that employs the inverted Wishart distribution. The percentage improvement shows that DHT model is better than the bivariate model in all cases except for the improvement in MAD for the prediction of interpurchase time in business-to-consumer dataset (-.52%).

[Table 4]

Among the four nested HB models, the IDHO model, where purchase monetary amount and interpurchase time were modeled independently with a common variance for all individuals, predicts worst. If we relax the homogenous assumption and allow the variance-parameters to be heterogeneous among customers (the IDHT model), we find that predictive performance generally improves in most cases (eight out of twelve percentage improvements of IDHT against IDHO are positive). The same is true when the independence assumption between purchase monetary amount and interpurchase time is relaxed (ten out of twelve percentage improvements of DHT versus DHO are positive). Both components, heterogeneity of variances across customers and allowing for dependencies within customers, are important elements of the model.

DISCUSSION

The superiority of the dependent, heterogeneous model (DHT) relative to the standard bivariate normal HB model is due to the effects of differential shrinkage. The DHT model uses the inverted gamma distributions for each the diagonal elements of the bivariate covariance matrix, which is more flexible than the inverted Wishart distribution used in the standard normal model. In addition, by factoring the joint distribution of interpurchase time and purchase monetary amount as the marginal distribution of interpurchase time, $[t]$, and the conditional distribution of purchase monetary amount, $[y|t]$, we can introduce a covariance term (η) that is not restricted along the real line. This allows for a more flexible parameterization of the covariance matrix while ensuring that, for any value of η , the covariance matrix is positive definite.

The effect of the proposed re-parameterization on the posterior distribution of (log) interpurchase time and purchase monetary amount is illustrated in Figure 4. Displayed are the 95% HPD ellipsoids of the posterior bivariate distribution for three customers from the business-to-consumer dataset. The bivariate distributions are evaluated at the posterior means of the model parameter for each customer. The solid line represents estimates from the proposed model, and the dashed line corresponds to the standard HB estimate using the Wishart distribution. The Wishart distribution is less sensitive to the presence of positive and/or negative covariation in interpurchase time and purchase monetary amount, whereas the proposed model more sensitive to covariation of these data.

[Figure 4]

We explore the implications of our model for customer scoring in Tables 5 and 6. Scoring involves a ranking of customers according to a scalar criterion such as customer value. A common score used in direct marketing is customer value, where value is measured in terms of

revenue per unit time (\$/time) (see Fader, Hardie, and Lee, 2005; Gupta, Hanssens, Hardie, Kahn, Kumar, Lin, Ravishanker, and Sriram, 2006). In our analysis we do not calculate a discounted value of a future income stream, and leave this extension of our model for future research. Once customers have been ranked by the score, a cutoff value (k) is applied, and customers for whom $\Pr((\$/\text{time}) > k)$ are selected for targeting activities (see also Bult and Wansbeek 1995).

The specification of $\Pr((\$/\text{time}) > k)$ is due to two factors. First, the value of k depends on the average expenditure for maintain the relationship with a customer. In practice, practitioners of relationship marketing would want to retain a customer that would bring net positive profits into the organization. Thus, k must be at least equal to, or greater than, average expenditure for maintain the relationship with a customer. Second, the probability of $(\$/\text{time}) > k$ must be high enough for the customer being targeted in the future. In our model, this translates to $\Pr(\$/\text{time} = \exp[y - t] > k)$ because y and t are measured logarithmically.

For the business-to-business dataset, a cutoff value of $k=.4$ was used to split customers into ten approximately equal-sized groups based on the probability that their revenue stream would exceed the cutoff. The value of .4 for k implies a target revenue of at least US \$544.5 per year (i.e., $(e^{.4}) \cdot 365$ days). For the business-to-consumer dataset, a cutoff value of $k=6.3$ was identified. The value of 6.3 for k implies a target revenue of at least NT \$198,769 per year (i.e., $(e^{6.3}) \cdot 365$ days). Using the cutoff, we:

1. Computed $\Pr(\$/\text{time} = \exp[y - t] > k)$ for each customer using the first half of the data (i.e. in-sample data) and 1000 draws from the posterior distribution of their model parameters. This calculation corresponds to the upper tail of a normal distribution.

2. Sorted customers into 10 probability groups (deciles). All customers for whom $\Pr(\$/\text{time} > k)$ was between .0 and .1 were grouped together in the first group, probabilities between .1 and .2 were used to form the second group, and so on, with probabilities between .9 and 1.0 used to form the tenth group.
3. Calculated the actual proportion of customers within each decile group with observed value ($\$/\text{time}$) greater than k in the second half of the data (i.e., out-sample data). That is, if a customer had an observed $\$/\text{time}$ greater than k in the second half of the data, this customer was assigned a value of one. Otherwise, the customer was assigned a value of zero. The actual proportion of customers with observed $\$/\text{time}$ greater than k was measured by dividing the number of customers assigned a value of one by the sample size of the decile group to which the customers was assigned.
4. Computed the mean absolute deviation (MAD) between the in-sample and out-of-sample probability values. The most accurate model is the one with predictive behavior most closely matching the in-sample calibration.

The left side of Table 5 begins with the ten decile groups and the number of customers assigned to each group for our proposed dependent-heterogeneous (DHT) model. Fifty four customers are assigned to the first group with lowest predicted probabilities (i.e., $\Pr(\$/\text{time} > k) < .1$). The average in-sample probability for these customers is .0706. Five of these customers had actual

$\$/\text{time}$ greater than the cutoff value $k=.4$ in the hold-out portion of the data, for a proportion of $5/54 = .0926$. Across the different models investigated, the cutoff value (k) can be seen to lead to different frequencies of customers in each decile group.

The predictive results reported in Tables 5 and 6 indicate that our proposed (DHT) model can lead to large improvements in predictive MAD. The gains are larger for the business-to-business dataset (Table 5), consistent with our detection of greater evidence of covariation in these data. The improvements in the DHT model are not concentrated in any one of the deciles. The predictive accuracy of the proposed DHT model is substantially better than the standard bivariate normal HB model for both datasets, and outperforms alternative, restricted forms of the proposed model.

[Tables 5 and 6]

The practical implications of our model are explored by comparing predicted cash flows for each of the models. As demonstrated earlier, erroneously assuming that purchase timing and amount are independent when they are actually positive will lead to an upward bias in predicted revenues. To illustrate the magnitude of this effect, we sorted customers from the business-to-business dataset on the basis of the $\Pr(\$/\text{time} > k)$ values for each of the five models, selected the top 20% as our target list (Koch, 1998), and calculated the actual purchase amount per year for each customer from the second half of the datasets. Thus, the number of respondents used in each analysis was the same, representing the top 20% most valued customers.

The average purchase amount is US\$1309 for DHT model and is higher than those obtained by the other competing models. The average purchase amount is US\$1305 for DHO model, US\$1204 for IDHT, and US\$1165 for bivariate model. These results indicate that the

choice of analysis models can lead to huge losses. For example, the average purchase amount for DHT is US\$144.10 greater than that obtained from bivariate model proposed by Ansari, Jedidi, and Jagpal (2000) and Ying, Feinberg, and Wedel (2006).

Since the whole dataset from which we drew the sample has five hundred thousand customers, this means that the top 20% customers (i.e., 100,000 customers) selected on the basis of the results of DHT model can bring about US\$14.4 million per year more than those top 20% customers selected on the basis of the results of bivariate model. Our estimates also indicate that the loss of assuming independence, as is commonly done in direct marketing, is close to US\$10 million based on results from the IDHT model ($(\$1309 - \$1204) * 10000$). Similar results were obtained for the business-to-consumer data set – compared to the DHT model, the losses due to employing the conventional bivariate model and the independent IDHT models are NT\$666 million and NT\$567 million, respectively. Thus, models assuming independence (e.g., IDHT) over-value customers and lead to a sub-optimal identification of high value customers.

CONCLUDING REMARKS

Our results indicate that the standard direct marketing assumption of independent behavioral components is not supported. We find evidence of compensating customer behavior in our data, where larger quantities of purchase are associated with longer inter-purchase times. We demonstrate that the presence of compensating behavior adversely affects the accuracy of customer scoring and the ability to identify the most profitable customers in models that assume independence (e.g., Thomas 2001; Reinartz and Kumar 2003; Gupta, Lehmann and Stuart 2004; Fader, Hardie and Lee 2005; Borle, Singh and Jain 2008).

This paper proposes a general approach to combining timing and quantity data into an overall measure of value that reflects a customer's income stream. A challenge in integrating inter-purchase time and purchase quantity data stems from the limited amount of data available for each customer, and the heterogeneity of both the mean level and covariation of these data. Standard approaches to building hierarchical Bayes estimates rely on the inverted Wishart distribution, which we find to be restrictive in measuring heterogeneous patterns of covariation. Our approach avoids use of the inverted Wishart distribution by building up the bivariate distribution from a marginal and a conditional distribution.

Our model is expected to work best with "shallow" datasets with less than ten observations per customer. For panel data with more observations per respondent, our simulation results indicate that allowing for heterogeneous covariation in purchase timing and amount using standard approach should work well. In either case, models that incorrectly assume independence between purchase timing and quantity will tend to over-value the value of a customer when compensating behavior is present – i.e., longer inter-purchase times compensating for larger purchase quantities. Thus, while the model and estimation algorithm presented are more computationally demanding than existing models, especially in large datasets, we show that the benefits are potentially large. The computational burden of our approach will decline as computing speeds continue to increase.

Covariance heterogeneity has not yet received much attention in the marketing literature, both in simple models of behavior and those that are more complicated. Simple examples include RFM measures, where estimates of recency, frequency and monetary value are independently calculated and aggregated to obtain an overall score for the customer. The presence of dependent effects leads to double-counting in the aggregate measure of some of the

elemental components. An example of a more complicated models of behavior include discrete choice models, where heterogeneous effects for the expected value of respondent utility are typically present, but not for elements of the covariance matrix. An exception is Ansari, Jedidi and Jagpal (2000). Our model generalizes the Ansari, Jedidi and Jagpal (2000) model by allowing for a less restrictive specification of the covariance matrix.

As researchers continue to develop new models of consumer behavior and explore new areas of application, we anticipate greater attention placed on properties of distributions and the use of non-standard methods. Our approach to modeling covariance heterogeneity can be extended to higher dimensional distributions in a straight-forward manner. For example, the joint distribution of any number of elements $a, b, c, \dots$ can be constructed by multiplying together appropriate marginal and conditional distributions, e.g., $[a][b|a][c|a,b]\dots$ etc.. We therefore anticipate application of this approach in other domains of research. An example is the development of hierarchical Bayes models for structural equation modeling (e.g., LISREL), which models the joint distribution of data using a series of conditional relationships such as $[c|a,b]$. In these models, the data $\{a,b,c\}$ is typically vector-valued, and the joint distribution can be of high dimension. The use of a Wishart distribution with a single degree-of-freedom parameter would be problematic in this case. We leave this application for future development.

There are also a number of avenues for future work in extending our model in direct marketing. First, our use of a normal distribution may not be appropriate in some direct marketing contexts and may need to be relaxed. The normal model implies that a symmetrical relationship which may be overly restrictive in some circumstances. Second, inter-purchase times and purchase quantities may have a lead-lag relationship which we do not explore in our

analysis. A dynamic model would more accurately reflect customer behavior in many product categories, and is a fruitful avenue for future research.

REFERENCES

- Allenby, Greg M. and James L. Ginter (1995), "Using extremes to design products and segment markets," *Journal of Marketing Research*, 32 (November), 392-403.
- Allenby, Greg M., Geraldine Fennell, Joel Huber, Thomas Eagle, Tim Gilbride, Dan Horsky, Jaehwan Kim, Peter Lenk, Rich Johnson, Elie Ofek, Bryan Orme, Thomas Otter and Joan Walker (2005) "Adjusting Choice Models to Better Predict Market Behavior," *Marketing Letters*, 16 (3-4), 197-208.
- Allenby, Greg M., Robert Leone, and Lichung Jen (1999), "A Dynamic Model of Purchase timing with Application to Direct Marketing," *Journal of the American Statistical Association*, 94, 365-374.
- Allenby, Greg M. and Peter E. Rossi (1999), Marketing Models of Consumer Heterogeneity, *Journal of Econometrics*, 89 (1/2), 57-78.
- Ansari, Asim, Kamel Jedidi, and Sharan Jagpal (2000), "A Hierarchical Bayesian Methodology for Treating Heterogeneity in Structural Equation Models," *Marketing Science*, 19 (4), 328-347.
- Ansari, Asim, Skander Essegiaier, and Rajeev Kohli (2000), "Internet Recommendation Systems," *Journal of Marketing Research*, 37 (August), 363-375.
- Baesens, Bare, Stijn Viaene, Dirk Van den Poel, Jan Vanthienen, and Guido Dedene (2002), "Bayesian Neural Networks Learning for Repeat Purchase Modeling in Direct Marketing," *European Journal of Operational Research*, 138 (1), 191-211.
- Blattberg, Robert. C., Gary Getz, and Jacquelyn Thomas (2001), *Customer Equity: Building and Managing Relationships as Valuable Assets*, Boston, Mass.: Harvard Business School Press.

- Boatwright, Peter, Sharad Borle, and Joseph B. Kadane (2003), "A Model of the Joint Distribution of Purchase Quantity and Timing," *Journal of the American Statistical Association*, 98 (463), 564-572.
- Borle, Sharad, Siddharth S. Singh and Dipak C. Jain (2008) "Customer Lifetime Measurement," *Management Science*, 54, 1, 100-112.
- Boulding, William, Richard Stelin, Michael Ehret, and Wesley J. Johnston (2005), "A Customer Relationship Management Roadmap: What is Known, Potential Pitfalls, and Where to Go," *Journal of Marketing*, 69 (4), 155-166.
- Bult, Jan Roelf and Tom Wansbeek (1995) "Optimal Selection for Direct Mail," *Marketing Science*, 14, 4, 378-394.
- Cao, Yong and Thomas S. Gruca (2005), "Reducing Adverse Selection through Customer Relationship Management," *Journal of Marketing*, 69 (4), 219-229.
- Colombo, Richard, and Weina Jiang. (1999), "A Stochastic RFM Model," *Journal of Interactive Marketing*, 13 (Summer), 2-12.
- Cullinan, G. J. (1977), *Picking Them by Their Batting Averages: Recency-Frequency-Monetary Method of Controlling Circulation*, New York, NY: Direct/Mail Marketing Association.
- Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2005), "RFM and CLV: Using Iso-CLV Curves for Customer Base Analysis," *Journal of Marketing Research*, 42 (November), 415-430.
- Gupta, Sunil, Dominique Hanssens, Bruce Hardie, William Kahn, V. Kumar, Nathaniel Lin, Nalini Ravishanker, and S. Sriram (2006), "Modeling Customer Lifetime Value," *Journal of Service Research*, 9 (2), 139-155.

- Hughes, Arthur M. (2005), *Strategic Database Marketing: The Masterplan for Starting and Managing a Profitable Customer-Based Marketing Program*, New York, NY: McGraw-Hill.
- Jacoby, Jacob, and David B. Kyner (1973), "Brand Loyalty vs. Repeat Purchasing Behavior," *Journal of Marketing Research*, 10 (February), 1-9.
- Jen, Lichung, Chou Chien-Heng, and Greg Allenby (2003), "A Bayesian Approach to Modeling Purchase Frequency," *Marketing Letters*, 14 (1), 5-20.
- Jen, Lichung, and Shih-Ju Wong (1998), "Incorporating Heterogeneity in Customer Valuation: An Empirical Study of Health Care Direct Marketing in Taiwan," *International Journal of Operations and Quantitative Management*, 4 (3), 217-228.
- Koch, Richard (1998), *The 80/20 Principle: The Secret of Achieving More with Less*, New York, NY: Doubleday.
- Lenser, John F. (2002), "Gimme an R. Gimme and F. Gimme an M." *Catalog Age*, 19 (7), 53-54.
- Leszinski, Ralf, Felix A. Weber, Roberto Raganoni, and Thomas Baumgartner (1995), "Profits in Your Backyard," *McKinsey Quarterly*, 4, 118-127.
- Miglautsch, John (2002), "Application of RFM Principles: What to Do with 1-1-1 Customers?" *Journal of Database Marketing*, 9 (4), 319-324.
- Morrison, Donald, and Schmittlein, David (1988), "Generalizing the NBD Model for Customer Purchases: What Are the Implications and Is It Worth the Effort?" *Journal of Business & Economic Statistics*, 6(2), 145-159.

- Newton, Michael A. and Adrian E. Raftery (1994), "Approximate Bayesian Inference with the Weighted Likelihood Bootstrap," *Journal of the Royal Statistical Society, Series B (Methodological)*, 56 (1), 3-48.
- Rainey, Brian (2001), "Transactional Modeling for Beginners," *Catalog Age*, 18 (4), 67.
- Reichheld, Frederick F. and Phil Schefter (2000), "E-loyalty: Your Secret Weapon on the Web," *Harvard Business Review*, 78 (4), 105-113.
- Reichheld, Frederick F., Robert G. Markey Jr., and Christopher Hopton (2000), "The Loyalty Effect – the Relationship between Loyalty and Profits," *European Business Journal*, 12(3), 134-139.
- Reinartz, Werner, and V. Kumar (2000), "On the Profitability of Long-Life Customers in a Noncontractual Setting: An Empirical Investigation and Implications for Marketing," *Journal of Marketing*, 64 (October), 17-35.
- Reinartz, Werner, and V. Kumar (2003), "The Impact of Customer Relationship Characteristics on Profitable Lifetime Duration," *Journal of Marketing*, 67 (1), 77-99.
- Rossi, Peter E. and Greg M Allenby (2003). "Bayesian Statistics and Marketing," *Marketing Science*, 22(3), 304-328.
- Rust, Roland T., Valarie A. Zeithaml and Katherine N. Lemon (2000) Driving Customer Equity: How Customer Lifetime Value is Reshaping Corporate Strategy, New York: The Free Press.
- Schmittlein, David C. and Robert A. Peterson (1994), "Customer Base Analysis: An Industrial Purchase Process Application," *Marketing Science*, 13 (Winter), 41-67.

- Sonnier, Garrett, Andrew Ainslie and Thomas Otter (2007) "Heterogeneity Distributions of Willingness-To-Pay in Choice Models," *Quantitative Marketing and Economics*, forthcoming.
- Stone, Bob and Ron Jacobs (2001), *Successful Direct Marketing Methods*, New York, NY: McGraw-Hill.
- Suman, Alicia O. (2002), "Gaining B-to-B RFM," *Target Marketing*, 25 (7), 51-53.
- Thomas, Jacquelyn S. (2001) "A Methodology for Linking Customer Acquisition to Customer Retention," *Journal of Marketing Research*, 38, 262-268.
- Venkatesan, Rajkumar and V. Kumar (2004), "A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy," *Journal of Marketing*, 68 (4), 106-125.
- Weinstein, Art (2002), "Customer Retention: A Usage Segmentation and Customer Value Approach," *Journal of Targeting Measurement & Analysis for Marketing*, 10 (3), 259-268.
- Ying, Yuanping, Fred Feinberg, and Michel Wedel (2006), "Leveraging Missing Ratings to Improve Online Recommendation Systems," *Journal of Marketing Research*, 43 (August), 355-365.
- Zehna, Peter W. (1966) "Invariance of Maximum Likelihood Estimators," *Annals of Mathematical Statistics*, 37, 744.

Table 1
Simulation Study Model Comparison
(Posterior Standard Deviations)

	<i>Simulated Dataset 1: With Equal Size of Variances of μ_y and μ_t</i>			<i>Simulated Dataset 2: With Unequal Size of Variances of μ_y and μ_t</i>		
	True Model	Bivariate Normal	Conditional Normal	True Model	Bivariate Normal	Conditional Normal
Hyper-Parameter Recovery Performance:						
μ_y	5.00	4.987 (.102)	4.990 (.013)	5.00	4.981 (.102)	4.979 (.022)
μ_t	2.00	2.000 (.043)	2.001 (.009)	2.00	1.993 (.041)	1.994 (.090)
σ_{yy}	2.00	1.972 (.066)	2.096 (.028)	10.00	10.220 (.332)	10.737 (.317)
σ_{tt}	2.00	1.889 (.049)	2.060 (.025)	2.00	1.988 (.050)	2.120 (.123)
σ_{yt}	1.00	.967 (.040)	1.034 (.020)	4.00	4.042 (.137)	4.280 (.149)
Individual-Parameter Recovery Performance:						
<i>MAD</i>						
μ_{y_i}		.279	.281		.548	.549
μ_{t_i}		.250	.248		.239	.241
σ_{yy_i}		.682	.481		2.639	1.909
σ_{tt_i}		.511	.360		.499	.345
σ_{yt_i}		.412	.312		1.035	.736
<i>RMSE</i>						
μ_{y_i}		.401	.393		.754	.761
μ_{t_i}		.355	.361		.332	.338
σ_{yy_i}		1.761	1.372		5.275	3.741
σ_{tt_i}		1.065	.679		1.046	.718
σ_{yt_i}		.841	.625		2.116	1.460
<i>Log Marginal Density</i>		-53.150	-42.380		-65.573	-49.098

Table 2
Parameter Estimates for B-B Data
Posterior Mean (Standard Deviation)

Parameters:	Regression	Bivariate Normal Hierarchical Bayes	Conditional Normal Hierarchical Bayes			
			Independent		Dependent	
			Homo- geneous	Hetero- geneous	Homo- geneous	Hetero- geneous
Marginal Distribution of Purchase Timing – Equation (4)						
$\upsilon_i \sim \text{Normal}(X_i'\gamma, \xi^2)$						
γ_1 : Northeast Area	4.361 (.030)	4.264 (.031)	4.260 (.030)	4.357 (.033)	4.261 (.029)	4.362 (.032)
γ_2 :Midwest Area	4.401 (.044)	4.295 (.043)	4.293 (.047)	4.385 (.045)	4.296 (.045)	4.381 (.046)
γ_3 :South Area	4.359 (.037)	4.250 (.039)	4.244 (.036)	4.342 (.039)	4.246 (.039)	4.351 (.038)
γ_4 :West Area	4.247 (.032)	4.168 (.034)	4.171 (.032)	4.257 (.034)	4.165 (.033)	4.256 (.034)
γ_5 :Insurance-related Firms	-.089 (.183)	-.089 (.180)	-.014 (.182)	-.081 (.194)	-.036 (.186)	-.078 (.203)
γ_6 :Medical Offices	-.356 (.091)	-.377 (.091)	-.373 (.092)	-.394 (.097)	-.375 (.090)	-.401 (.099)
γ_7 :Attorneys	.037 (.071)	.033 (.075)	.042 (.071)	.022 (.070)	.040 (.075)	.018 (.072)
ξ^2	--	--	.365 (.017)	.413 (.018)	.362 (.019)	.414 (.018)
$\zeta_i^2 \sim \text{Inverse Gamma}(\kappa_2, \delta_2)$						
κ_2	--	--	--	1.178 (.013)	--	1.178 (.013)
δ_2	--	--	--	1.006 (.026)	--	1.004 (.026)
Conditional Distribution of Purchase Quantity – Equation (5)						
$\mu_i \sim \text{Normal}(X_i'\beta, \zeta^2)$						
β_1 : Northeast Area	4.183 (.021)	4.177 (.023)	4.176 (.022)	4.189 (.021)	3.994 (.029)	3.981 (.034)
β_2 :Midwest Area	4.184 (.031)	4.182 (.030)	4.182 (.033)	4.186 (.034)	3.998 (.041)	3.993 (.044)
β_3 :South Area	4.181 (.026)	4.171 (.025)	4.168 (.025)	4.194 (.028)	3.976 (.032)	3.982 (.040)
β_4 :West Area	4.106 (.023)	4.108 (.022)	4.106 (.024)	4.119 (.024)	3.922 (.031)	3.892 (.032)
β_5 :Insurance-related Firms	-.019 (.129)	-.006 (.132)	.025 (.125)	.073 (.133)	.073 (.143)	.062 (.135)
β_6 :Medical Offices	.067 (.064)	.068 (.057)	.070 (.058)	.080 (.066)	.117 (.064)	.083 (.078)
β_7 :Attorneys	.084 (.050)	.081 (.049)	.085 (.051)	.077 (.044)	.112 (.055)	.076 (.057)
ζ^2	--	--	.159 (.009)	.167 (.009)	.100 (.007)	.130 (.008)
$\eta_i \sim \text{Normal}(\varphi, \varpi^2)$						
φ	--	--	--	--	.043 (.005)	.048 (.006)
ϖ^2	--	--	--	--	.010 (.0004)	.011 (.0005)
$\sigma_i^2 \sim \text{Inverse Gamma}(\kappa_1, \delta_1)$						
κ_1	--	--	--	1.467 (.018)	--	1.385 (.016)
δ_1	--	--	--	1.274 (.028)	--	1.177 (.027)
The Components of the Variance-Covariance Matrix for Bivariate Normal Model						
σ_{yy}	--	.00213 (.00008)				
σ_{tt}	--	.00120 (.00004)				
σ_{yt}	--	-.00008 (.00001)				
Log Marginal Density		-42062	-43607	-42597	-43157	-41967

Table 3
Parameter Estimates for B-C Data
Posterior Means (Standard Deviation)

Parameters:	Regression	Bivariate Normal Hierarchical Bayes	Conditional Normal Hierarchical Bayes			
			Independent		Dependent	
			Homo- geneous	Hetero- geneous	Homo- geneous	Hetero- geneous
Marginal Distribution of Purchase Timing – Equation (4)						
$v_i \sim \text{Normal}(X_i' \gamma, \xi^2)$						
γ_1 : Female	3.031 (.103)	2.996 (.093)	2.955 (.096)	2.936 (.095)	2.950 (.096)	2.930 (.103)
γ_2 : Male	3.055 (.099)	3.021 (.084)	2.976 (.093)	2.962 (.092)	2.968 (.088)	2.958 (.097)
γ_3 : North Area	-.088 (.103)	-.090 (.089)	-.048 (.096)	-.053 (.096)	-.042 (.095)	-.051 (.102)
γ_4 : West Area	-.036 (.114)	-.046 (.102)	-.001 (.108)	-.0006 (.1103)	.008 (.104)	.004 (.110)
γ_5 : South Area	-.085 (.119)	-.089 (.104)	-.044 (.115)	-.049 (.110)	-.037 (.114)	-.042 (.115)
γ_6 : East Area	-.194 (.108)	-.189 (.095)	-.147 (.099)	-.149 (.096)	-.143 (.100)	-.140 (.104)
ξ^2	--	--	.499 (.023)	.485 (.025)	.499 (.024)	.484 (.023)
$\zeta_i^2 \sim \text{Inverse Gamma}(\kappa_2, \delta_2)$						
κ_2	--	--	--	1.309 (.014)	--	1.309 (.015)
δ_2	--	--	--	1.031 (.028)	--	1.031 (.026)
Conditional Distribution of Purchase Quantity – Equation (5)						
$\mu_i \sim \text{Normal}(X_i' \beta, \zeta^2)$						
β_1 : Female	8.755 (.063)	8.699 (.062)	8.712 (.064)	8.679 (.061)	8.755 (.062)	8.672 (.061)
β_2 : Male	8.680 (.060)	8.632 (.059)	8.646 (.061)	8.618 (.059)	8.696 (.058)	8.605 (.057)
β_3 : North Area	.161 (.063)	.214 (.062)	.211 (.064)	.210 (.062)	.178 (.061)	.165 (.060)
β_4 : West Area	.219 (.070)	.270 (.068)	.270 (.072)	.263 (.067)	.215 (.068)	.206 (.066)
β_5 : South Area	.045 (.073)	.105 (.075)	.107 (.074)	.112 (.069)	.046 (.071)	.046 (.073)
β_6 : East Area	.081 (.066)	.135 (.064)	.132 (.066)	.136 (.065)	.091 (.065)	.090 (.062)
ζ^2	--	--	.199 (.009)	.196 (.008)	.144 (.009)	.115 (.007)
$\eta_i \sim \text{Normal}(\varphi, \varpi^2)$						
φ	--	--	--	--	-.0001 (.005)	.024 (.005)
ϖ^2	--	--	--	--	.012 (.001)	.017 (.001)
$\sigma_i^2 \sim \text{Inverse Gamma}(\kappa_1, \delta_1)$						
κ_1	--	--	--	2.000 (.026)	--	1.826 (.026)
δ_1	--	--	--	2.005 (.046)	--	1.732 (.041)
The Components of the Variance-Covariance Matrix for Bivariate Normal Model						
σ_{yy}	--	.00600 (.00023)				
σ_{tt}	--	.00315 (.00011)				
σ_{yt}	--	-.00017 (.00002)				
Log Marginal Density		-84726	-89972	-88768	-88110	-84681

Table 4
Predictive Performance
(Percentage Improvement of DHT Model)

Model	<u>Business-to-Business</u>		<u>Business-to-Consumer</u>	
	RMSE	MAD	RMSE	MAD
Regression				
Log Monetary Amount (y)	.871 (6.51%)	.786 (7.76%)	.730 (16.13%)	.647 (20.91%)
Log Time (t)	1.178 (5.25%)	1.036 (5.07%)	1.172 (6.95%)	1.018 (8.45%)
Log Customer Value (y/t)	1.342 (6.12%)	1.161 (5.90%)	1.471 (12.80%)	1.275 (14.25%)
Bivariate Normal Hierarchical Bayes				
Log Monetary Amount (y)	.875 (6.97%)	.739 (1.33%)	.700 (11.30%)	.547 (2.22%)
Log Time (t)	1.265 (13.00%)	1.030 (4.40%)	1.173 (7.07%)	.934 (-.52%)
Log Customer Value (y/t)	1.439 (13.81%)	1.124 (2.54%)	1.447 (10.95%)	1.127 (1.02%)
Independent & Homogeneous (IDHO)				
Log Monetary Amount (y)	.826 (1.00%)	.741 (1.52%)	.642 (2.09%)	.550 (2.68%)
Log Time (t)	1.154 (3.08%)	1.025 (3.90%)	1.093 (-.26%)	.936 (-.27%)
Log Customer Value (y/t)	1.268 (.30%)	1.098 (.18%)	1.321 (1.30%)	1.133 (1.53%)
Independent & Heterogeneous (IDHT)				
Log Monetary Amount (y)	.821 (.43%)	.735 (.77%)	.634 (.79%)	.540 (.88%)
Log Time (t)	1.117 (-.17%)	.985 (-.18%)	1.096 (.01%)	.939 (.01%)
Log Customer Value (y/t)	1.271 (.51%)	1.101 (.43%)	1.309 (.38%)	1.121 (.45%)
Dependent & Homogeneous (DHO)				
Log Monetary Amount (y)	.821 (.43%)	.734 (.62%)	.638 (1.50%)	.545 (1.87%)
Log Time (t)	1.153 (3.07%)	1.025 (3.88%)	1.092 (-.31%)	.936 (-.35%)
Log Customer Value (y/t)	1.281 (1.31%)	1.112 (1.48%)	1.313 (.71%)	1.126 (.88%)
Dependent & Heterogeneous (DHT)				
Log Monetary Amount (y)	.818	.730	.629	.535
Log Time (t)	1.119	.986	1.096	.939
Log Customer Value (y/t)	1.264	1.096	1.304	1.116

Table 5
Predictive Performance of Customer Scoring in B-B Data
Pr(\$/Time > k)
(Percentage Improvement of DHT Model)

Pr(\$/time>k)	DHT			DHO			IDHT			IDHO			Bivariate		
	Fre- quency ^a	In- sample ^b	Out- sample ^c	Fre- quency	In- sample	Out- sample	Fre- quency	In- sample	Out- sample	Fre- quency	In- sample	Out- sample	Fre- quency	In- sample	Out- sample
.0 - .1	54	.071	.093	24	.076	.125	120	.061	.150	43	.069	.163	58	.067	.138
.1 - .2	103	.147	.185	82	.153	.134	127	.151	.252	113	.154	.168	133	.154	.181
.2 - .3	142	.258	.247	115	.252	.226	112	.246	.286	136	.250	.235	134	.250	.291
.3 - .4	99	.345	.323	124	.349	.282	71	.348	.380	105	.345	.305	88	.351	.341
.4 - .5	86	.454	.395	107	.445	.365	67	.446	.403	76	.452	.461	61	.451	.393
.5 - .6	72	.549	.458	71	.548	.394	54	.548	.444	57	.547	.456	47	.540	.511
.6 - .7	52	.652	.596	64	.648	.469	49	.643	.531	55	.648	.418	67	.648	.448
.7 - .8	58	.747	.638	59	.750	.695	49	.750	.633	47	.750	.617	35	.751	.486
.8 - .9	42	.849	.691	48	.852	.688	35	.853	.571	41	.853	.610	39	.843	.641
.9 - 1.0	39	.954	.872	53	.957	.811	63	.962	.825	74	.970	.824	85	.968	.800
MAD			.065			.094			.106			.101			.107
(Improve Rate)						(45%)			(63%)			(57%)			(65%)

Note: a: number of observations whose probability of $\text{Pr}(\$/\text{time} = \exp[y - t] > k)$ falls within the range specified in the most left-hand side column. We set k to be .4, which implies that the target revenue for each customer is US \$544.5 per year (i.e., $(e^4) \cdot 365$ days).

b: The average $\text{Pr}(\$/\text{time} = \exp[y - t] > k)$ for all the observations assigned to a decile group, using the first half of the data (i.e. within a time span of one and a half years).

c: The actual proportion of customers within each decile group with observed value (\$/time) greater than k in the second half of the data (i.e. within a time span of one and a half years).

Table 6
Predictive Performance of Customer Scoring in B-C Data
Pr(\$/Time > k)
(Percentage Improvement of DHT Model)

Pr(\$/time>k)	DHT			DHO			IDHT			IDHO			Bivariate		
	Fre- quency ^a	In- sample ^b	Out- sample ^c	Fre- quency	In- sample	Out- sample	Fre- quency	In- sample	Out- sample	Fre- quency	In- sample	Out- sample	Fre- quency	In- sample	Out- sample
.0 - .1	159	.026	.302	154	.027	.286	191	.015	.293	192	.014	.297	200	.012	.305
.1 - .2	37	.147	.243	31	.157	.387	30	.145	.400	29	.145	.448	26	.142	.423
.2 - .3	32	.242	.563	35	.238	.429	20	.253	.600	18	.251	.444	16	.241	.500
.3 - .4	19	.358	.421	27	.353	.444	13	.351	.539	15	.342	.667	14	.357	.643
.4 - .5	25	.456	.640	21	.461	.571	19	.452	.632	16	.453	.563	14	.463	.714
.5 - .6	24	.548	.583	28	.554	.679	17	.555	.529	18	.556	.611	17	.559	.471
.6 - .7	33	.653	.788	38	.651	.816	18	.650	.833	18	.653	.778	16	.660	.750
.7 - .8	53	.750	.836	72	.756	.847	22	.750	.864	28	.751	.857	23	.753	.783
.8 - .9	94	.854	.848	91	.856	.857	46	.858	.783	40	.862	.825	34	.862	.853
.9 - 1.0	165	.961	.855	144	.965	.833	265	.981	.853	267	.981	.843	281	.983	.847
MAD			.131			.140			.177			.168			.172
(Improve Rate)						(7%)			(35%)			(28%)			(32%)

Note: a: number of observations whose probability of $\text{Pr}(\$/\text{time} = \exp[y - t] > k)$ falls within the range specified in the most left-hand side column. We set k to be 6.3, which implies that the target revenue for each customer is NT\$198,769 per year (i.e., $(e^{6.3}) \cdot 365$ days).

b: The average $\text{Pr}(\$/\text{time} = \exp[y - t] > k)$ for all the observations assigned to a decile group, using the first half of the data (i.e. within a time span of one and a half years).

c: The actual proportion of customers within each decile group with observed value (\$/time) greater than k in the second half of the data (i.e. within a time span of one and a half years).

Figure 1
COMPARISON OF MODEL FIT IN THE PRESENCE OF UNEQUAL VARIANCES.

Original Density

Figure 2
SEQUENCE OF MCMC DRAWS FOR η_i AND σ_i^2

Figure 3
THE DISTRIBUTION OF CORRELATION COEFFICIENTS (ρ_i) IN THE B-B AND B-C DATASETS

Figure 4
COMPARISON OF 95% HPD ELLIPSOIDS FOR SELECTED CUSTOMERS
(— : DHT MODEL; ---- : BIVARIATE NORMAL HB MODEL)

THE IMPORTANCE OF MODELING TEMPORAL DEPENDENCE OF TIMING AND QUANTITY IN DIRECT MARKETING

Lichung Jen, Chien-Heng Chou, and Greg M. Allenby

WEB APPENDIX

THE GIBBS SAMPLER FOR CONDITIONAL NORMAL HIERARCHICAL BAYES MODEL OF PURCHASE QUANTITY AND INTERPURCHASE TIME

Denote y_{ij} as the j^{th} log purchase amount of the i^{th} customer, and t_{ij} as the j^{th} log interpurchase time of the i^{th} customer, where $i=1,2,\dots, N$ and $j=1,2,\dots, n_i$. Assume that the conditional distribution of y_{ij} given t_{ij} is normally distributed as

$$f(y_{ij} | t_{ij}) \sim N(\mu_i + t_{ij}\eta_i, \sigma_i^2),$$

and that t_{ij} is also normally distributed as

$$f(t_{ij}) \sim N(\nu_i, \varsigma_i^2).$$

Then, the joint distribution of Y_{ij} and t_{ij} is

$$f(y_{ij}, t_{ij}) \sim N\left(\begin{bmatrix} \mu_i + \nu_i\eta_i \\ \nu_i \end{bmatrix}, \begin{bmatrix} \eta_i^2\varsigma_i^2 + \sigma_i^2 & \eta_i\varsigma_i^2 \\ \eta_i\varsigma_i^2 & \varsigma_i^2 \end{bmatrix}\right).$$

The prior and the conditional posterior distributions of all parameters are summarized below.

1. The conditional posterior distribution of $\mu_i | y_{ij}, t_{ij}, x_i, \beta, \sigma_i, \eta_i, \zeta$ is

$$\text{Normal} \left[\left(\frac{n_i}{2\sigma_i^2} + \frac{1}{2\zeta^2} \right)^{-1} \left(\frac{\sum_{j=1}^{n_i} (y_{ij} - t_{ij}\eta_i)}{\sigma_i^2} + \frac{(x_i'\beta)}{\zeta^2} \right), \left(\frac{n_i}{\sigma_i^2} + \frac{1}{\zeta^2} \right)^{-1} \right].$$

The prior of μ_i is:

$$\text{Normal}(x_i'\beta, \zeta^2).$$

1.1 The conditional posterior distribution of $\beta | \mu_i, x_i, \zeta, b_0, V_{b_0}$ is

$$\text{Normal} \left(\left(V_{b_0}^{-1} + \zeta^{-2} \sum x_i x_i' \right)^{-1} \left(V_{b_0}^{-1} b_0 + \zeta^{-2} \sum \mu_i x_i' \right), \left(V_{b_0}^{-1} + \zeta^{-2} \sum x_i x_i' \right) \right).$$

The prior of β is

$$\text{Normal}(b_0, V_{b_0}),$$

where $b_0=4$ and

$$V_{b_0} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}.$$

1.2 The conditional posterior distribution of $\zeta^2 | \mu_i, x_i, \beta, s_0, S_0$ is

$$\text{Inverse Gamma} \left(s_0 + \frac{N}{2}, \left(\frac{1}{2} \sum_{i=1}^N (\mu_i - x_i'\beta)^2 + S_0^{-1} \right)^{-1} \right).$$

The prior of ζ^2 is Inverse Gamma(s_0, S_0), where $s_0=2$ and $S_0=0.25$.

2. The conditional posterior distribution of $\eta_i \mid y_{ij}, t_{ij}, \mu_i, \sigma_i, \varphi, \varpi$ is

$$\text{Normal} \left(\left(\frac{\left(\sum_{j=1}^{n_i} t_{ij}^2 \right)}{2\sigma_i^2} + \frac{1}{2\varpi^2} \right)^{-1} \left(\frac{\sum_{j=1}^{n_i} t_{ij} (y_{ij} - \mu_i)}{\sigma_i^2} + \frac{\varphi}{\varpi^2} \right), \left(\frac{\left(\sum_{j=1}^{n_i} t_{ij}^2 \right)}{\sigma_i^2} + \frac{1}{\varpi^2} \right)^{-1} \right).$$

The prior of η_i is $\text{Normal}(\varphi, \varpi^2)$.

2.1 The conditional posterior distribution of $\varphi \mid \eta_i, \varpi, f_0, F_0^2$ is

$$\text{Normal} \left(\left(\frac{N}{\varpi^2} + \frac{1}{F_0^2} \right)^{-1} \left(\frac{\sum_{i=1}^N \eta_i}{\varpi^2} + \frac{f_0}{F_0^2} \right), \left(\frac{N}{\varpi^2} + \frac{1}{F_0^2} \right)^{-1} \right).$$

The prior of φ is $\text{Normal}(f_0, F_0^2)$, where $f_0=0$ and $F_0^2=4$.

2.2 The conditional posterior distribution of $\varpi^2 \mid \eta_i, \varphi, w_0, W_0^2$ is

$$\text{Inverse Gamma} \left(w_0 + \frac{N}{2}, \left(\frac{I}{2} \sum_{i=1}^N (\eta_i - \varphi)^2 + (W_0^2)^{-I} \right)^{-I} \right).$$

The prior of ϖ^2 is $\text{Inverse Gamma}(w_0, W_0^2)$, where $w_0=2$ and $W_0^2=0.5$.

3. The conditional posterior distribution of $\sigma_i^2|y_{ij}, t_{ij}, \mu_i, \eta, \kappa_1, \delta_1$ is

$$\text{Inverse Gamma} \left(\kappa_1 + \frac{n_i}{2} \quad , \quad \left(\frac{1}{2} \sum_{j=1}^{n_i} \left(y_{ij} - (\mu_i + t_{ij} \eta_i) \right)^2 + \delta_1^{-1} \right)^{-1} \right).$$

The prior of σ_i^2 is Inverse Gamma(κ_1, δ_1).

3.1 The conditional posterior distribution of $\kappa_1, \sigma_i, \delta_1$ is proportional to

$$\prod_{i=1}^N \frac{\left(\frac{I}{\delta_I \sigma_i^2} \right)^{\kappa_I} e^{\frac{-I}{\delta_I \sigma_i^2}}}{\Gamma(\kappa_I)},$$

which is the product of N continuous Poisson distributions with parameters of $1/(\delta_1 \sigma_i^2)$. The prior of κ_1 is Uniform(10).

3.2 The conditional posterior distribution of $\delta_1|\sigma_i, \kappa_1, d_0, D_0$ is

$$\text{Inverse Gamma} \left(d_0 + N\kappa_1 \quad , \quad \left(\sum_{i=1}^N \sigma_i^{-2} + D_0^{-1} \right)^{-1} \right).$$

The prior of δ_1 is Inverse Gamma(d_0, D_0), where $d_0=2$ and $D_0=2$.

4. The conditional posterior distribution of $v_i | t_{ij}, \zeta_i, \gamma, \xi$ is

$$\text{Normal} \left[\left(\frac{n_i}{2\zeta_i^2} + \frac{1}{2\xi^2} \right)^{-1} \left(\frac{\sum_{j=1}^{n_i} t_{ij}}{\zeta_i^2} + \frac{(x_i' \gamma)}{\xi^2} \right), \left(\frac{n_i}{\zeta_i^2} + \frac{1}{\xi^2} \right)^{-1} \right].$$

The prior of t_{ij} is $\text{Normal}(x_i' \gamma, \xi_i^2)$.

4.1 The conditional posterior distribution of $\gamma | v_i, x_i, \xi, g_0, V_{g_0}$ is

$$\text{Normal} \left(\left(V_{g_0}^{-1} + \xi^{-2} \sum x_i x_i' \right)^{-1} \left(V_{g_0}^{-1} g_0 + \xi^{-2} \sum v_i x_i' \right), \left(V_{g_0}^{-1} + \xi^{-2} \sum x_i x_i' \right)^{-1} \right).$$

The prior of γ is Multivariate Normal(g_0, V_{g_0}), where $g_0=4$ and

$$V_{g_0} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}.$$

4.2 The conditional posterior distribution of $\xi^2 | v_i, x_i, \gamma, s_0, S_0$ is

$$\text{Inverse Gamma} \left(s_0 + \frac{N}{2}, \left(\frac{1}{2} \sum_{i=1}^N (v_i - x_i' \gamma)^2 + S_0^{-1} \right)^{-1} \right).$$

The prior of ξ^2 is Inverse Gamma(s_0, S_0), where $s_0=2$ and $S_0=0.25$.

5. The conditional posterior distribution of $\zeta_i^2 | t_{ij}, v_i, \kappa_2, \delta_2$ is

$$\text{Inverse Gamma} \left(\kappa_2 + \frac{n_i}{2}, \left(\frac{1}{2} \sum_{j=1}^{n_i} (t_{ij} - v_i)^2 + \delta_2^{-1} \right)^{-1} \right).$$

The prior of ζ_i^2 is Inverse Gamma(κ_2, δ_2).

5.1 The conditional posterior distribution of $\kappa_2 | \zeta_i, \delta_2$ is proportional to

$$\prod_{i=1}^N \frac{\left(\frac{1}{\delta_2 \zeta_i^2} \right)^{\kappa_2} e^{-\frac{1}{\delta_2 \zeta_i^2}}}{\Gamma(\kappa_2)},$$

which is the product of N continuous Poisson distributions with parameters of

$$\left(\frac{1}{\delta_2 \zeta_i^2} \right).$$

The prior of κ_2 is Uniform(M_{κ_0}), where $M_{\kappa_0}=10$.

5.2 The conditional posterior distribution of $\delta_2 | \zeta_i, \kappa_2, d_0, D_0$ is

$$\text{Inverse Gamma} \left(d_0 + N\kappa_2, \left(\sum_{i=1}^N \zeta_i^{-2} + D_0^{-1} \right)^{-1} \right).$$

The prior of δ_2 is Inverse Gamma(d_0, D_0), where $d_0=2$ and $D_0=2$.