Greg M. Allenby is the Cullman Professor of Marketing, Fisher College of Business, Ohio State University, Columbus, Ohio 43210 (Email: allenby.1@osu.edu). Robert P. Leone is the Berry Chair in Marketing Fisher College of Business, Ohio State University, Columbus, Ohio 43210 (Email: leone.7@osu.edu). Lichung Jen is Associate Professor of Business, College of Management, National Taiwan University, Taipei, Taiwan, ROC (Email: lichung@haydn.ntu.edu.tw). The authors thank Jim Pedrick for providing helpful comments and access to the data, and Peter Rossi and the referees for many helpful comments.
A DYNAMIC MODEL OF PURCHASE TIMING
WITH APPLICATION TO DIRECT MARKETING

Abstract
Predicting changes in individual customer behavior is an important element for success in any direct marketing activity. In this paper we develop a hierarchical Bayes model of customer inter-purchase times which is based on the generalized gamma distribution. The model allows for both cross-sectional and temporal heterogeneity, and can be used to predict when and if a specific customer will likely increase their time between purchases. This prediction can be used managerially as a signal for the firm to employ some type of intervention to keep that customer.

Keywords: hierarchical Bayes, generalized gamma distribution, panel data
1. INTRODUCTION

Successful direct marketing initiatives require firms to predict the behavior of specific individuals. At major investment brokerage firms such as Merrill Lynch, Fidelity and Schwab, for example, five to ten percent of customers switch firms each year, taking with them billions of dollars in assets and close to a hundred million dollars in revenue. A critical problem faced by these investment firms, and for many direct marketing companies, is predicting when specific individuals are likely to stop being regular customers and enter periods of relative inactivity. With such prediction, firms can aggressively intervene when they feel a customer is likely to move from a frequent to a less frequent level of purchase behavior due to the fear that they are taking their business to a competitor.

In recent years the field of direct marketing has grown considerably, largely due to the reduced costs associated with collecting and storing customer records and the availability of distribution channels (e.g. the internet, fax machines and firms such as Federal Express) that provide direct access to the customer. However, despite the ready availability of customer records, direct marketers currently operate in an environment characterized by data sets with severe limitations. For example, while information might be available to a firm about the date and price paid for an item, there is typically no information on the availability of goods from other manufacturers or whether a consumer had considered these other goods when making their purchase. Therefore it is often not possible to estimate constructs such as a consumer’s utility function which is assumed to drive the purchase decision. Instead, firms have been forced to use statistical models of
purchase data that, by necessity, ignore competitive effects and unobserved customer behavior (e.g. purchases of competitive products).

In this paper we develop a dynamic model of purchase timing that provides an early indication of when specific customers are likely to move from an active state to a less active state. The model has three distinct features: (1) it employs a generalized gamma distribution which we find useful for modeling inter-purchase times; (2) random-effects are introduced through a hierarchical Bayes structure which yields individual-level estimates of key model parameters and functions of parameters; and (3) temporal variation in an individual’s expected inter-purchase time is permitted by introducing covariates that change through time. From a direct marketing perspective there are two major advantages to this model. First, the hierarchical Bayes random-effects specification, coupled with the temporal component of the model, allows for a customer specific baseline rate of interaction with the firm that can fluctuate over time. Second, and more important, the covariates permit prediction of when this rate is likely to change.

In the next section we discuss the data and provide descriptive statistics. In section 3 we introduce the model and discuss the estimation algorithm. Results are provided in section 4 where model predictions are compared to simpler methods of generating individual-level forecasts of future inter-purchase times. Concluding remarks are offered in section 5.

2. THE DATA

The data are provided by a major investment brokerage firm that buys and sells financial instruments such as stocks, bonds and mutual funds for their customers.
Revenues are generated as a percentage of the trade amount. In consultation with senior managers and analysts from the firm, it was decided that inter-purchase times (in this case, inter-trading times) should be measured in terms of calendar weeks. There were three reasons for this. First, managers at the firm felt that for the purpose of decision making, it was not necessary to react to changes in inter-purchase times in units less than a week. They did not see any advantage in a model of daily (versus weekly) purchase timing. Second, we found many investors would sell stocks on one day and then buy other stocks on the next day with the proceeds of that sale. Working with weekly data helped to minimize these a financial trades which spanned more than one day but were most likely related to a single decision to make a trade through the firm. Third, the use of weekly data avoids calendar effects (weekends and holidays) in which, for example, a trade on a Friday and the subsequent Monday would appear to have an inter-purchase time of three days which does not reflect the fact that markets were closed on Saturday and Sunday. Therefore, all trades within a particular week were aggregated and treated as one trade.

The data span a total of six years beginning in January, 1991, and includes account-level information such as the trade date, amount of transaction ($) and whether the account was originally opened for personal investing versus a retirement/planned savings account in which investments occur on a regular time schedule. Our analysis focuses on a random sample of 251 customers who opened new accounts in the first six months of 1991 for the purpose of personal investing, producing a total of 4573 inter-purchase times. Complete purchase histories are available for each of these 251 customers. A sample of 325 customers (6452 inter-purchase times) who opened new account in the last six months of 1991 was also selected for predictive testing of the
models. Difference in the sample sizes is due to more customers opening accounts in the later part of 1991.

Figure 1 displays the average inter-purchase time for each of the 251 customers in the estimation sample. The average times range from 1.17 weeks to 104 weeks, indicating a very heterogeneous population of customers. Looking at figure 1 it is clear that some customers trade nearly every week, while others are much less involved in the management of their dollars held by the firm. This fact illustrates the importance of allowing for cross-sectional heterogeneity in customer inter-purchase times in any model of buying behavior.

To illustrate the modeling problem faced, figure 2 provides time-series plots of the inter-purchase times for three customers in the sample. The plots show large and abrupt changes in the inter-purchase times sometimes occur. The large values of inter-purchase time correspond to periods of inactivity that the firm would like to prevent. This could be done by various means such as having a broker contact the particular customer and asking if their needs are being met, or by mailing out some investment literature. Regardless of the means, the success and efficiency of such programs depend critically on obtaining sufficient advanced warning of a customer’s move to an inactive state.

In summary, a simple analysis of the data reveals the need for a probability density with positive support that allows for cross-sectional heterogeneity and temporal dynamics. In addition, in order for direct marketers to manager their customers, the model must allow for inferences about the behavior of individual customers and be capable of generating disaggregate forecasts of their behavior. In the next section we introduce a model which addresses each of these issues.
3. THE MODEL

We start by assuming that \( t_{ij} \), the \( j \)th inter-purchase time for customer \( i \), is distributed generalized gamma (GG):

\[
t_{ij} \sim GG(\alpha, \lambda_i, \gamma) = \frac{\gamma \alpha^\gamma \lambda_i^\alpha}{\Gamma(\alpha)} t_{ij}^{\alpha-1} e^{-\frac{\gamma t_{ij}}{\lambda_i}}
\] (1)

The parameters \( \alpha \) and \( \gamma \) establish the shape of the distribution (see Johnson, Kotz and Balakrishnan 1994, p.388), and \( \lambda_i \) is a scale parameter which varies across customers.

The expected value of the generalized gamma distribution is equal to \( \frac{\Gamma(\alpha+1/\gamma)}{\Gamma(\alpha)} \lambda_i \).

The generalized gamma is a distribution which is related to other distributions as follows. If \( x_i \sim \text{exponential}(\lambda) \) then \( y = \sum_{i=1}^{\alpha} x_i \) is distributed \( \text{gamma}(\alpha, \lambda) \). Raising \( y \) to the \( 1/\gamma \) power, \( z = \left( \sum_{i=1}^{\alpha} x_i \right)^{1/\gamma} \) results in a random variate from the generalized gamma distribution, \( GG(\alpha, \lambda, \gamma) \). Special cases of the generalized gamma distribution are the exponential (\( GG(1, \lambda, 1) \)), Erlang (\( GG(2, \lambda, 1) \)), gamma (\( GG(\alpha, \lambda, 1) \)), and Weibull (\( GG(1, \lambda, \gamma) \)) distributions. The distribution is defined on the positive real numbers (consistent with using time as the dependent variable) and very flexible in the variety of shapes and implied hazard functions it can represent. For example, non-monotonic hazards (i.e. increasing then decreasing) routinely occur for \( \gamma < 1 \).

The generalized gamma distribution has been used in the analysis of lifetime data with limited success. As discussed by Lawless (1982), this was originally due to
identification problems with different values of \( \alpha, \lambda \) and \( \gamma \) leading to very similar distributions, thus requiring the use of large datasets to obtain accurate parameter estimates. Reparameterization of the density and improvements in estimation methods have resulted in greater use of the density, although a survey by the authors of the *Current Index to Statistics* reveals relatively few applications that use the generalized gamma distribution. Some recent articles include its application to reliability analysis (Agarwal, 1996), material stress (Pham and Almhana, 1996) and the duration of economic events (Jaggia, 1991). In this paper we find that the generalized gamma distribution offers a flexible basis for building models of inter-purchase timing.

### 3.1 Cross-sectional Heterogeneity

Heterogeneity across customers is modeled by assuming an inverse generalized gamma distribution (IGG) for \( \lambda_i \):

\[
\lambda_i \sim \text{IGG}(\nu, \theta; \gamma) = \frac{\gamma}{\Gamma(\nu) \theta^\nu} \lambda_i^{\nu-1} e^{-\left(\frac{\lambda_i}{\theta}\right)^\nu}
\]

(2)

The mean of the inverse generalized gamma is equal to \( \frac{\Gamma(\nu-1/\gamma)}{\Gamma(\nu)\theta} \). This distribution is also very flexible and, as shown in the appendix, results in tractable conditional distributions to execute MCMC estimation (Gelfand and Smith 1990). The model specified by equations (1) and (2) conditions on \( \gamma \). In the analysis presented below, we identify \( \gamma \) through a specification search in which we calculate the posterior odds of alternative models (e.g. \( \gamma=0.5, 1.0, 2.0, \ldots \)). We complete the model specification for \( t_{ij} \) by assuming prior distributions for \( \nu \) to be IGG(a,b,\( \gamma \)) where a and b are chosen to have
minimal influence on the conditional distributions (i.e. a diffuse prior), and a uniform prior on $\theta$.

In contrast to our model, other models of purchase incidence have often assumed a Poisson distribution for the number of purchases ($z$) in a given time period ($T$), and a gamma distribution of heterogeneity of the purchase rate $\lambda_i^*$ (see Ehrenberg 1988 and Schmittlein, Morrison and Colombo 1987). The purchase rate is the reciprocal of the inter-purchase time (i.e. $\lambda_i^* = 1/\lambda_i$). In theory, there is no difference between a model which assumes exponential inter-purchase times with an inverse gamma distribution of heterogeneity, and one which assumes a Poisson purchase rate with a gamma distribution of heterogeneity. Both yield mathematically equivalent inferences and predictions. An advantage of our hierarchical Bayes model of purchase timing is that it generalizes these standard models because the exponential and gamma distributions are special cases of the generalized gamma distribution.

Finite mixture models (Heckman and Singer 1984, Kamakura and Russell 1989) have been proposed as a method to minimize the impact of distributional assumptions of heterogeneity in the analysis of duration data. The advantage of using a discrete distribution of heterogeneity is also partly due to the fact that the likelihood is easily evaluated as the sum of the likelihood over the mass points. For a sufficient number of mass points, any distribution can be approximated to a high degree of accuracy. However, there is a growing body of literature in marketing (Allenby and Ginter 1995, Allenby and Rossi 1998) which demonstrates that a finite mixture model with a small number of mass points does not adequately represent the full extent of heterogeneity. A disadvantage of
the finite mixture model is that estimates of $\lambda_i$ are constrained to lie within the convex hull of the point masses. As a result, the finite mixture model leads to estimates of individual effects which are much less heterogeneous than those obtained from continuous mixing distributions. We therefore use a continuous mixing distribution in our analysis.

3.2 Temporal Dynamics

Variation in expected inter-purchase time for a particular customer is modeled by relating $\lambda_i$ to covariates. This can be easily accomplished in two ways. First, $\lambda_i$ can be related to the covariates through a multiplicative model:

$$\lambda_{ij} = \lambda_i \delta_1^{x_{ij1}} \delta_2^{x_{ij2}} \cdots \delta_k^{x_{ijk}}$$  \hspace{1cm} (3)

where $x_{ij}$ are time varying covariates and $\lambda_i$ is distributed inverse generalized gamma across consumers. The advantage of this formulation is that it can be easily estimated via MCMC methods (see appendix). Further, we have:

$$\lambda_{ij} = \lambda_{i0} \delta_1^{x_{ij1} - E[x_{ij1}]} \delta_2^{x_{ij2} - E[x_{ij2}]} \cdots \delta_k^{x_{ijk} - E[x_{ijk}]} = \left( \frac{\lambda_{i0}}{\delta_1^{E[x_{ij1}]} \delta_2^{E[x_{ij2}]} \cdots \delta_k^{E[x_{ijk}]} } \right) \delta_1^{x_{ij1}} \delta_2^{x_{ij2}} \cdots \delta_k^{x_{ijk}}$$  \hspace{1cm} (4)

where the term in brackets is equal to $\lambda_i$ in equation (3). If values of the covariates are equal through time they have no effect on the model, while temporal variation in the covariates ($x$) result in variation in $\lambda_{ij}$. If the covariates are measured in logarithmic form, the coefficients $\delta$ can be interpreted as the percentage change in the expected inter-purchase time for a percentage change in the covariate. This is true because the expected inter-purchase time is proportional to $\lambda$ in (1) above.

A second approach to incorporating temporal dynamics is with a generalized gamma component mixture model:
\[ t_{ij} \sim \sum_k \phi_{ijk} GG(\alpha_k, \lambda_{ik}, \gamma_k) \] (5)

where \( \phi_{ijk} \) is related to the covariates \( (x) \), and \( \lambda_{ik} \) is assumed distributed \( IGG(\nu_k, \theta_k, \gamma_k) \).

This model allows for a more abrupt change in the predicted inter-purchase times than equation (4) because, for example, the parameters of the generalized gamma distribution can reflect very different expected inter-purchase times. Time varying covariates can be related to the mass points \( \phi_{ijk} \) through a variety of link functions such as a binomial, ordered, or multinomial probit model. The specification used in our specific analysis is described next.

3.3 A Model for Purchase Timing

The goal of our analysis is to obtain an early indication of when customers are likely to become inactive. We use three lagged inter-purchase times, in logarithmic form, as covariates \( (x) \) in our analysis, and allow for three component densities \( (k=3) \). The use of additional component densities did not result in substantively better model fit or predictions. Thirty two percent of the observations in our data take on the value of one, corresponding to a super-active state of trading in which customers are buying and selling stocks each week. The first component density reflects these individuals. The remaining two components are used to model active and inactive trading behavior, with covariates used to provide early predictions of when a customer is likely to move from the second (active) component to the third (inactive) component. We impose the ordering \( \theta_1 > \theta_2 > \theta_3 \) to achieve statistical identification of these states, and specify the component probabilities as:
\[ \phi_{ij1} = 1 - \Phi(\beta_{01i}) \quad \text{(super-active state)} \quad (6) \]

\[ \phi_{ij2} = \Phi(\beta_{01i}) (1 - \Phi(x_{ij} \beta_{2i})) \quad \text{(active state)} \quad (7) \]

\[ \phi_{ij3} = \Phi(\beta_{01i}) \Phi(x_{ij} \beta_{2i}) \quad \text{(inactive state)} \quad (8) \]

where \( \Phi \) denotes the standard normal cdf and \( x_{ij} = (\log(t_{ij-1}), \log(t_{ij-2}), \log(t_{ij-3}))' \) is a vector of three lagged log inter-purchase times. Other aspects of the trade (e.g. the amount of the transaction ($)) and market conditions (e.g. behavior of the stock market) were investigated but were not found to be statistically significant predictors of customer behavior. In addition, we restrict the value of \( \alpha_k \) to values less than 50 through truncated prior distributions to avoid having the first component density collapse around the integer value of one for the super-active trades. Finally, we allow the parameters \( \beta_{i} = (\beta_{01i}, \beta_{2i}) \) to be heterogeneously distributed across customers by a random-effects specification:

\[ \beta_{i} \sim \text{Normal}(\bar{\beta}, V) \quad (9) \]

This model captures important distinctions between components that correspond to super-active, active and inactive customer states. In equation (6), each customer has a unique stationary probability of being in a super-active state. This probability is not related to their past behavior and is constant across observations (j). In our analysis of the data we find that customers often remain in this state for many weeks, and therefore the lagged inter-purchase times (equal to one in this instance) have no predictive value.
because the covariates are constant during these spells. Lagged inter-purchase times are used to allow for temporal variation in the probability that a customer is either active (equation 7) or inactive (equation 8). The strength of this relationship is allowed to vary across customers according to the random-effects specification (equation 9), implying that for some customers their past behavior is a strong predictor of their future behavior while for others it is not.

An alternative approach to modeling our data is with a flexible specification of the hazard function (see Jain and Vlccassim 1991, Gonul and Srinivasan 1993) that allows for long periods of customer inactivity. Our approach differs from a proportional hazard model in two respects. First, the heterogeneity parameter \( \lambda_i \) serves to rescale the time axis while maintaining the basic shape of the density. This differs from proportional hazard models in which heterogeneity is typically used to shift the baseline hazard function (not the density). More important, the introduction of covariates into our model facilitates the prediction of when customers are more likely to change from an active to an inactive state. This differs from the use of thick-tailed hazard functions, which can capture long periods of customer inactivity but do not necessarily predict when these events occur – the critical question when deciding whether to contact a customer.

4. ANALYSIS OF INVESTMENT DATA

In this section we compare various models to illustrate (1) the advantage of using the generalized gamma relative to the standard gamma density, and (2) the advantage of incorporating temporal dynamics relative to simpler approaches. In our analysis of the data we find that temporal dynamics based on the mixture model (equation 5) yield a
better fit to the data relative to a model that relates the parameter $\lambda_i$ to covariates through equation (4). As stated earlier, the temporal variation in inter-purchase times displayed in figure 2 requires a model which allows for abrupt shifts in behavior which the mixture model can easily accommodate. Therefore, the remainder of the discussion will focus on our model which incorporates temporal dynamics through a mixture model.

4.1 Parameter Estimates

Table 1 reports parameter estimates for four models. The first and second models employ a gamma density for $t_{ij}$ and inverse gamma heterogeneity ($\gamma=1$). The first model does not allow for temporal dynamics, while the second model does. The third and fourth models use the (inverse) generalized gamma distribution instead of the (inverse) gamma. The third model does not allow for temporal dynamics while the fourth model does. Also reported is the log marginal density of the data that provides a measure of model fit. This statistic was computed using the importance sampling estimate as suggested by Newton and Raftery (1994, p.21). Values of $\gamma_k$ were obtained via a grid search procedure that minimized the marginal density of the data. The estimates reported in the table indicate improvement in model fit resulting from use of the generalized gamma density and the lagged inter-purchase times.

Figure 3 displays the component likelihood distributions reported in table 1 for the gamma and generalized gamma models (with temporal dynamics). The component distributions are evaluated at the posterior mean of $\lambda_{ik}$. The distribution for different individuals will differ through the rescaling of the time (horizontal) axis induced by the heterogeneity in $\lambda$. The expected inter-purchase time for the generalized gamma model
(model 3), evaluated at the mean value of $\lambda_{ik}$, is equal to $\frac{\Gamma(\alpha_k + 1/\gamma_k) \Gamma(\nu_k - 1/\gamma_k)}{\Gamma(\alpha_k) \Gamma(\nu_k) \theta_k}$.

Using this formula, the parameters of the generalized gamma model yields expected inter-purchase times of 1.00, 1.52 and 21.93 weeks for each of the three components, corresponding to super-active, active and inactive customer states. In contrast, expected inter-purchase times for components of the model using the gamma likelihood are equal to 1.00, 3.00 and 19.14 weeks. The use of the generalized gamma distribution therefore results in different component densities than those obtained with the gamma distribution.

Hazard functions for the three component densities of generalized gamma model are provided in figure 4. While all three of the hazard functions are monotonically increasing, the hazard for the first two components have much greater range than the hazard for the third component which is nearly constant over a wide range. This indicates that if an observation belongs to the inactive state (third component), there is little that can be predicted about the probable timing of the customer’s order. It is therefore highly desirable for the firm to have a large customer base with inter-purchase times distributed according to the density of the first two mixing components.

Figure 5 displays the distribution of heterogeneity of $\lambda_i$ for the second and third mixing components of the generalized gamma model. The first component is massed tightly at the value of 0.02 and is not displayed. The 90 percent HPD for the second component is (0.25, 0.37) and (3.16, 16.46) for the third component. Figure 5 indicates that consumer heterogeneity is an important characteristic of the data, which is consistent with the distribution of average inter-purchase times plotted in figure 1.
Parameter estimates which determine the mass points ($\phi$) for the component densities (equations 6, 7 and 8) are reported at the bottom of table 1. The means of the lagged inter-purchase time coefficients are all positive, indicating that on average, as previous inter-purchase times increase (relative to an individual’s average inter-purchase time), the likelihood the next observation belongs to the third (inactive) component increases. The model therefore provides an early prediction of the likelihood that a consumer is moving into this undesirable state. However, the variance of the random effects indicates that early predictions are possible for some individuals but not all.

The primary difference between the models is with the use of the generalized gamma versus the gamma distribution. The generalized gamma distribution results in more of the observations classified into the third component (0.437 versus 0.410 when the gamma distribution is used) and slightly larger slope coefficients for the mass points. This difference, coupled with the greater expected inter-purchase time of the third component when using the generalized gamma distribution (see figure 3), results in a more aggressive model with better fit to the data. The extent to which this improvement in in-sample fit translates to improved out-of-sample predictions is discussed next.

4.2 Predictive Performance

A holdout sample of 325 customers (6452 observations) was used for predictive testing of the models. In our analysis, we use all but the last observation for each individual to estimate expected inter-purchase times for each individual and the probability ($\phi$) that the last observation belongs to each of the two components. To simulate the decision making conditions which we expect to encounter in practice, we condition our
predictive analysis on the point estimates of the parameters reported in table 1. That is, we do not update estimates of these hyper-parameters, but we do use the data to generate the individual-level estimates of expected inter-purchase times and the component probability. This results in an algorithm which is very fast with posterior statistics for the entire sample available within a few minutes which would be very appealing for practical applications.

Predictive accuracy is measured in two ways. The first is the mean absolute deviation between the predicted inter-purchase time and the actual time. For the component mixture models, the predicted value is a weighted average of the expected value of the three component densities for each individual. The weights (ϕ) are a function of the customer’s lagged last three inter-purchase times for models that incorporate temporal dynamics. In addition to the models discussed above, three other models are included for comparison. These are (1) a multiplicative model which incorporates the three lagged inter-purchase times into the model through equation (3); (2) a moving average model which uses the last three inter-purchase times to predict the next inter-purchase time; and (3) a simple average model which uses the entire purchase history to form the average. In general, the predictive error is large across all models, reflecting a noisy forecasting environment where there exists substantial variation in a customer’s inter-purchase times. Based on the mean absolute deviation statistic, the generalized gamma model with temporal dynamics is the most predictively accurate.

Also reported in table 2 is the correlation between the mass of the third (inactive) component and the actual inter-purchase time. This statistic is reported because it is more closely related to how a firm would employ the model in practice. That is, a firm would
use the model to identify customers with a high likelihood of being in an inactive state by examining the predicted mass of the third component ($\phi_3$). We find that the predicted value of $\phi_3$ is more closely correlated to the actual inter-purchase time when temporal dynamics are allowed in the model ($r=0.134$) than when they are not ($r=0.101$) or when a gamma distribution is used in place of the generalized gamma distribution ($r=0.099$ and $0.094$).

5. DISCUSSION

Marketers are interested in identifying when the behavior of their customers change, especially direct marketers who maintain individual customer data bases and have the ability to contact the customer. This is often a difficult task because of the limited information available to them. Some customers are very regular in the buying behavior (e.g. constant, predictable inter-purchase times) while others are more erratic and less predictable. Establishing benchmarks for this latter group requires individual-specific norms and measures so that statistical outliers can be identified.

In this paper we develop a random-effects generalized gamma component mixture model in which the probability that a particular observation belongs to a specific component is related to the past behavior of the customer. In applying this model to an investment database from a major brokerage firm, we are able to identify three components corresponding to super-active, active and inactive states. Estimating the model with MCMC methods results in individual-level estimates of summary statistics such as the expected inter-purchase times for each component and the probability that a specific observation belongs to each component.
In practice, firms could use the model to identify customers who are likely to move from an active state to a less active state in several ways. First, probability estimates for membership of the next inter-purchase time in a particular component can be periodically calculated and monitored over time. When this probability increases appreciably, it may be worthwhile for the firm to contact the customer and assess their level of satisfaction with the firm or products that have been provided. Also, with limited resources the probabilities could be ordered to create a priority structure of who to contact first.

Second, it is sometimes desirable to determine when it is advantageous to contact a customer who has not recently placed an order. That is, the customer’s data has a relatively large right-censored spell and it is of interest to determine whether this is an indication that the customer has moved to a less active state. For the current observation this can be assessed by using the hazard function (figure 4) to calculate the likelihood of a purchase in the near future given the right-censored spell and determining whether inactivity in this period is inconsistent with model. In addition, the predicted effect of a period of inactivity on the next observation can be made by calculating the conditional expectation of the inter-purchase time given the right-censored spell. For our three component model this is a weighted sum of the conditional expectations for the three densities. This estimate could then be used in equations (6), (7) and (8) to determine a revised probability that the next inter-purchase time is a member of the less active (as opposed to more active) state.

Predicting whether a customer will be active in the future is dependent on the frequency and regularity of their past purchase behavior, and whether their past behavior has historically been a good predictor of their future behavior. Our model is well suited
for this type of assessment because of its random-effects specification for both purchase
timing (equation 2) and component mixing (equations 6-8) portions. The random-effects
specification for purchase timing allows customers to differ in the expected inter-purchase
times associated with an active versus less active state. Conditional on these inter-
purchase times, a second random-effects model allows different customer’s mass points
($\phi_{ijk}$) to exhibit different degrees of dependence on lagged inter-purchase times.
Therefore the model captures key elements needed for establishing individual customer
benchmarks which identify changes in behavior.

In summary, the generalized gamma distribution provides a useful foundation for
building models of inter-purchase times and other behavior described by positive real
numbers. It is a flexible distribution which nests commonly used densities such as the
gamma and Weibull, and is conjugate to the inverse generalized gamma for random-effects
specification. The resulting random-effects models are easy to estimate using MCMC
methods, and can be expanded in various ways to include covariates of interest. While the
paper present one specific application of the model, there are numerous opportunities for
applying the model to direct marketing problems.
Appendix
Estimation Algorithms

MCMC estimation of the model defined by equations 1 and 2 proceeds by recursively generating draws from each of the following densities:

**Generate** $\lambda_i$ (one at a time for each customer):

$$
\pi(\lambda_i | \{t_{ij}, j=1,\ldots,n_i\}, n_i, \alpha, \nu, \gamma) = IGG(n_i \alpha + \nu, [\sum t_{ij}^\nu + \theta^\gamma]^{-1/\gamma}, \gamma) \quad (A.1)
$$

where: $t_{ij} = j^{th}$ inter-purchase times for customer $i$

$n_i =$ number of observations for customer $i$.

**Generate** $\alpha$:

$$
\pi(\alpha | \{\lambda_i\}, \{t_{ij}\}, \gamma) \propto \prod_{i=1}^{N} \prod_{j=1}^{n_i} \frac{\gamma}{\Gamma(\alpha) \lambda_{ij}^{\alpha \gamma} t_{ij}^{\alpha \gamma-1}} e^{-\frac{\gamma}{\lambda_{ij}}} \quad (A.2)
$$

**Generate** $\nu$:

$$
\pi(\nu | \{\lambda_i\}, \theta, \gamma) \propto \prod_{i=1}^{N} \frac{\gamma}{\Gamma(\nu) \theta^\gamma} \lambda_{ij}^{-\nu \gamma} e^{-\frac{\gamma}{\lambda_{ij}}} \quad (A.3)
$$

**Generate** $\theta$:

$$
\pi(\theta | \{\lambda_i\}, \nu, \gamma) = IGG(N\nu + a_0, [\sum \lambda_{ij}^{-\gamma} + b_0^{-\gamma}]^{-1/\gamma}, \gamma) \quad (A.4)
$$
where $a_0$ and $b_0$ are parameters from a prior distribution on $\theta$: $\pi(\theta) = \text{IGG}(a_0, b_0, \gamma)$. In our application we set $a_0 = 10$ and $b_0 = 10$. Given the large size of our dataset, these prior parameters exert minimal influence on the posterior.

For the multiplicative model (equation 3) with two covariates ($x_1$ and $x_2$), the conditional densities become:

\[
\pi(\lambda_i | \{t_{ij}, j=1,\ldots,n_i\}, n_i, x_{1ij}, x_{2ij}, \alpha, \nu, \gamma, \delta_1, \delta_2) = \text{IGG}(n_i \alpha + \nu, \Sigma_j (t_{ij} / \delta_1^{x_{1ij}} \delta_2^{x_{2ij}}) \gamma + \theta \gamma)^{-1/\gamma}, \gamma)
\]

\[
\pi(\delta_1 | \{\lambda_i\}, \{t_{ij}\}, \gamma, \delta_2, x_1, x_2) = \text{Gamma}(\alpha, \gamma) \prod_{i=1}^{N} \prod_{j=1}^{n_i} \frac{k_{ij}}{(\delta_1^{\lambda_{1ij}})^{\alpha \gamma}} e^{-\left(\frac{t_{ij}/\lambda_i \delta_1^{x_{1ij}}}{\delta_1^{\lambda_{1ij}}}\right)^{\gamma}} \quad (k_{ij} \text{ is the integrating constant})
\]

\[
\pi(\delta_2 | \{\lambda_i\}, \{t_{ij}\}, \gamma, \delta_1, x_1, x_2) = \text{Gamma}(\alpha, \gamma) \prod_{i=1}^{N} \prod_{j=1}^{n_i} \frac{k_{ij}}{(\delta_2^{\lambda_{2ij}})^{\alpha \gamma}} e^{-\left(\frac{t_{ij}/\lambda_i \delta_2^{x_{2ij}}}{\delta_2^{\lambda_{2ij}}}\right)^{\gamma}} \quad (k_{ij} \text{ is the integrating constant})
\]

\[
\pi(\alpha | \{\lambda_i\}, \{t_{ij}\}, \{y_{ij}\}, \gamma, \delta_1, \delta_2, x_1, x_2) = \text{Gamma}(\alpha, \gamma) \prod_{i=1}^{N} \prod_{j=1}^{n_i} \frac{\gamma}{\Gamma(\alpha) (\lambda_i \delta_1^{x_{1ij}} \delta_2^{x_{2ij}})^{\alpha \gamma}} e^{-\left(\frac{t_{ij}/\lambda_i \delta_1^{x_{1ij}} \delta_2^{x_{2ij}}}{\lambda_i \delta_1^{x_{1ij}} \delta_2^{x_{2ij}}}\right)^{\gamma}}
\]
and the conditional densities for \( \nu \) and \( \theta \) remains the same as in A.3 and A.4. Generating random draws form (A.6) - (A.8) is straightforward because they are univariate distributions and therefore it is easy to numerically evaluate the cumulative density and transform uniform (0,1) random draws.

The mixture model defined by equations (1), (2), (5), (6), (7) and (8) can be estimated MCMC methods using the standard trick (see for example Carlin and Chib, 1994) of introducing a latent random variable (\( s_{ijk} \)) which is used to index which of the \( k \) components each of the \( ij \) observations belong. On each iteration of the procedure, only those observations currently assigned to a component are used to arrive at \( \{ \lambda_{ik} \}, \alpha_k, \nu_k \) and \( \theta_k \). The assignment of observations to components is straightforward if \( \{ \phi_k \} \), the mass points of the components, are known: \( \Pr(s_{ijk} = k') \propto \phi_k f(t_{ij} | \alpha_k', \lambda_{ik}', \gamma_k') \) where \( \phi_k' \) is determined from equation (6). We impose the ordering \( \theta_1 > \theta_2 > \ldots > \theta_k \) to achieve model identification.

Draws of \( \beta_i \) are obtained from a Metropolis-Hastings algorithm with a random walk chain (see Chib and Greenberg, 1995). That is, denote \( \beta_i^{(p)} \) be the previous draw for \( \beta_i \). The next draw \( \beta_i^{(n)} \) is then given by:

\[
\beta_i^{(n)} = \beta_i^{(p)} + \Delta \beta
\]

where \( \Delta \beta \) is a draw from the density Normal(0,0.05). The choice for parameters of this density yields an acceptance rate of about 50\%. The acceptance probability is given by:

\[
P(\text{accept}) = \min \left[ \frac{\exp[-\frac{1}{2} (\beta_i^{(n)} - \bar{\beta})' V^{-1} (\beta_i^{(n)} - \bar{\beta}) \prod_j \phi_{ijk} (\beta_i^{(n)})]}{\exp[-\frac{1}{2} (\beta_i^{(p)} - \bar{\beta})' V^{-1} (\beta_i^{(p)} - \bar{\beta}) \prod_j \phi_{ijk} (\beta_i^{(p)})]} \right]
\]

(A.9)
where $\phi_{ik}(\beta_i)$ denotes the value obtained by evaluating equations 6-8 at $\beta_i$. The remaining conditional distributions are given by:

$$
\pi(\bar{\beta} | \{\beta_i\}, V) = \text{Normal}(\sum_{i=1}^{N} \beta_i / N, V / N) \tag{A.10}
$$

and

$$
\pi(V | \{\beta_i\}, \bar{\beta}) = \text{Inverted Wishart}(\sum_{i=1}^{N} (\beta_i - \bar{\beta})(\beta_i - \bar{\beta})' + G, N + g) \tag{A.11}
$$

where $G$ and $g$ are prior parameters which are set to $G=15I$, $g=15$. 
References


<table>
<thead>
<tr>
<th>Parameter Estimates and Posterior Standard Deviations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Likelihood</strong></td>
</tr>
<tr>
<td><strong>Heterogeneity</strong></td>
</tr>
<tr>
<td><strong>Temporal Dynamics</strong></td>
</tr>
<tr>
<td><strong>First Component</strong></td>
</tr>
<tr>
<td>$\gamma_1$</td>
</tr>
<tr>
<td>$\alpha_1$</td>
</tr>
<tr>
<td>$\nu_1$</td>
</tr>
<tr>
<td>$\theta_1$</td>
</tr>
<tr>
<td>mass</td>
</tr>
<tr>
<td><strong>Second Component</strong></td>
</tr>
<tr>
<td>$\gamma_2$</td>
</tr>
<tr>
<td>$\alpha_2$</td>
</tr>
<tr>
<td>$\nu_2$</td>
</tr>
<tr>
<td>$\theta_2$</td>
</tr>
<tr>
<td>mass</td>
</tr>
<tr>
<td><strong>Third Component</strong></td>
</tr>
<tr>
<td>$\gamma_3$</td>
</tr>
<tr>
<td>$\alpha_3$</td>
</tr>
<tr>
<td>$\nu_3$</td>
</tr>
<tr>
<td>$\theta_3$</td>
</tr>
<tr>
<td>mass</td>
</tr>
<tr>
<td><strong>Mass Point Parameters</strong></td>
</tr>
<tr>
<td>Intercepts:</td>
</tr>
<tr>
<td>$\beta_{01}$</td>
</tr>
<tr>
<td>$\beta_{02}$</td>
</tr>
<tr>
<td>Lagged Log Inter-Purchase Times:</td>
</tr>
<tr>
<td>t-1</td>
</tr>
<tr>
<td>t-2</td>
</tr>
<tr>
<td>t-3</td>
</tr>
<tr>
<td><strong>Variance of Random Effects</strong></td>
</tr>
<tr>
<td>Intercepts:</td>
</tr>
<tr>
<td>$\beta_{01}$</td>
</tr>
<tr>
<td>$\beta_{02}$</td>
</tr>
<tr>
<td>Lagged Log Inter-Purchase Times:</td>
</tr>
<tr>
<td>t-1</td>
</tr>
<tr>
<td>t-2</td>
</tr>
<tr>
<td>t-3</td>
</tr>
<tr>
<td><strong>Log Marginal Density</strong></td>
</tr>
</tbody>
</table>
Table 2. Predictive Performance of Various Models

<table>
<thead>
<tr>
<th>Likelihood</th>
<th>Heterogeneity</th>
<th>Temporal Dynamics</th>
<th>Mean Absolute Deviation</th>
<th>Correlation $(t, \phi_3)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gamma</td>
<td>Inverse Gamma</td>
<td>No</td>
<td>16.48</td>
<td>0.094</td>
</tr>
<tr>
<td>Gamma</td>
<td>Inverse Gamma</td>
<td>Yes</td>
<td>16.46</td>
<td>0.099</td>
</tr>
<tr>
<td>Generalized Gamma</td>
<td>Inverse Generalized Gamma</td>
<td>No</td>
<td>16.43</td>
<td>0.101</td>
</tr>
<tr>
<td>Generalized Gamma</td>
<td>Inverse Generalized Gamma</td>
<td>Yes</td>
<td>16.27</td>
<td>0.134</td>
</tr>
<tr>
<td>Multiplicative Model$^1$</td>
<td>Yes</td>
<td></td>
<td>16.80</td>
<td></td>
</tr>
<tr>
<td>Moving Average</td>
<td>Yes</td>
<td></td>
<td>18.29</td>
<td></td>
</tr>
<tr>
<td>Simple Average</td>
<td>No</td>
<td></td>
<td>17.97</td>
<td></td>
</tr>
</tbody>
</table>

$^1$ The multiplicative model (equation 3) uses the generalized gamma likelihood and inverse generalized gamma distribution for heterogeneity.
Figure 1. Distribution of Customer Average Inter-Purchase Times. Horizontal axis measured in calendar weeks.
Figure 2. Time Series Plots of Inter-Purchase Times. The horizontal axis is measured in calendar weeks. Peaks in the plots correspond to periods of customer inactivity that firms hope to avoid.
Figure 3. Component Distributions. The solid line corresponds to the model using the gamma likelihood and inverse gamma distribution for heterogeneity. The dotted line is for the model with generalized gamma likelihood and inverse generalized gamma distribution for heterogeneity. The first component corresponds to the super-active trading state, the second component corresponds to the active trading state and the third component is the inactive state. The horizontal axis is measured in weeks.
Figure 4. Component Hazard Functions. Hazards are for the generalized gamma model with temporal dynamics evaluated at the posterior mean of the distribution of heterogeneity for $\lambda_{ik}$. 
Figure 5. Distribution of Heterogeneity. Inverse generalized gamma distribution of $\lambda_{ik}$ for active (component 2) and inactive (component 3) states. Horizontal axis is in weeks.