

A Model of the Joint Distribution of Purchase Quantity and Timing

Peter Boatwright, Sharad Borle, and Joseph B. Kadane*

* Authors are listed in alphabetical order. Peter Boatwright is Assistant Professor of Marketing, GSIA, Carnegie Mellon University, Pittsburgh, PA 15213. Sharad Borle is Assistant Professor of Marketing, Jones Graduate School of Management, Rice University, Houston, TX 77251 and Joseph B. Kadane is Leonard J. Savage University Professor of Statistics and Social Sciences, Carnegie Mellon University, Pittsburgh, PA 15213. His research is supported by National Science Foundation grant DMS-9801401. This work was completed while Sharad Borle was a PhD student at GSIA, Carnegie Mellon University.

ABSTRACT

Prediction of purchase timing and quantity decisions of a household is an important element for success of any retailer. This is especially so for an online retailer, as the traditional brick-and-mortar retailer would be more concerned with total sales. A number of statistical models have been developed in the marketing literature to aid traditional retailers in predicting sales and analyzing the impact of various marketing activities on sales. However, there are two important differences between traditional retail outlets and the increasingly important online retail/delivery companies, differences that prevent these firms from using models developed for the traditional retailers: 1) the profits of the online retailer/delivery company depend on purchase frequency and on purchase quantity, while the profits of traditional retailers are simply tied to total sales, and 2) customers in the tails of the frequency distribution are more important to the delivery company than to the retail outlet. Both of these differences are due to the fact that the delivery companies incur a delivery cost for each sale, while customers themselves travel to retail outlets when buying from traditional retailers. These differences in costs translate directly into needs that a model must address. For a model intended to be useful to online retailers the dependent variable should be a bivariate distribution of frequency and quantity, and frequency distribution must accurately represent consumers in the tails.

In this article we develop such a model and apply it to predicting the consumer's joint decision of *when* to shop and *how much* to spend at the store. Our approach is to model the marginal distribution of purchase timing and the distribution of purchase quantity conditional on purchase timing. We propose a hierarchical Bayes model that disentangles the weekly and daily components of the purchase timing. The daily component has a dependence on the weekly component thereby accounting for strong observed periodicity in the data. For the purchase times, we use the Conway-Maxwell-Poisson distribution, which we find useful to fit data in the tail regions (extremely frequent and infrequent purchasers).

KEY WORDS: Conway-Maxwell-Poisson distribution; Hierarchical Bayes; Online retailers/delivery companies.

1. INTRODUCTION

When are my customers going to purchase and *how much* are they going to spend on their purchase are two important questions that any retailer would like to answer. These questions gain added importance in the case of an online retailer/delivery service where the pick, pack and delivery costs are substantial, in some cases more than half of the operating costs (O'Neill and Chu, 2001). Traditional retailers generally are less concerned with the separate *when* and *how much* effects, mostly focusing on total sales. Online retailers/delivery service, however, have a very different cost structure from that of traditional retailers (“The Milkman Returns – With Much More”, *Wall Street Journal*, December 5, 1999). Because delivery costs are substantial, the online retailer cannot simply focus on total sales but must be concerned with purchase frequency and purchase quantity. For example, promotions that induce frequent small quantity purchases may actually reduce profits at an online grocer while increasing profits at a traditional brick-and-mortar grocer.

Knowledge of a household's timing and amount of purchase decisions will help the online store to actively target these individual households and customize its offerings (Rossi, McCulloch, and Allenby, 1996), either to induce them to shop more frequently or in some cases less frequently. The ‘when’ and ‘how much’ decisions of the household also have important implications for the pricing structure for the delivery occasion and delivery volume. Typically these vary from a pay-per-use to fixed monthly amounts with a variety of combinations of these two.

Due to the large quantity of data on grocery purchases that have been available, a large number of sales models have been developed in the marketing literature. Such models have been utilized to set price, promotion and advertising policies (Jedidi, Mela and Gupta, 1999) and much of their intended contribution has been to give practitioners better tools for understanding their markets (Bucklin and Gupta, 1999); for example Boatwright, McCulloch & Rossi (1999) use their model to aid manufacturers allocating large trade promotion budgets across retailers. Due to the cost structure of traditional retailers, most of the existing models focus exclusively on retailer profits as a function of total sales, i.e. the models

are developed to predict sales. So although there are a number of models of grocery sales that have been developed, the online retailer would need a different type of model from a majority of these, needing a model that would jointly predict purchase frequency and purchase quantity rather than simply a model of total sales. Almost all of the few existing models in the literature for the joint distribution of purchase frequency and purchase quantity condition on known purchases elsewhere in the store (Bell and Lattin, 1998; Bucklin and Gupta, 1991; Gupta, 1988) or treat the frequency and quantity decisions as independent (Van Praag and Vermeulen, 1993; Meghir and Robin, 1992). A recent model of unconditional purchase frequency (Allenby, Leone, and Jen, 1999) utilized data from an investment brokerage to model the time between transactions for each customer. An alternative method of accommodating purchase frequency and quantity is to aggregate sales over long time frames such as months or quarters (Boatwright and Nunes, 2001), although aggregate models do not offer insight into the separate effects of frequency and quantity.

The analyst of grocery data faces two additional challenges, a partial data problem and customer heterogeneity. The first problem is partly addressed in the case of regular Brick-and-Mortar grocery stores by the availability of syndicated data, although the prevalence of “frequent shopper data,” data from customer panels tracked by individual retailers, has again highlighted this problem for traditional retailers as well. Since there is not yet a readily available syndicated source of data for online grocers, an individual online retailer must rely on purchase records of individuals at its own business. Because the model is designed for the retailer, the model must be designed to utilize information available to the retailer. Inference on customer purchase frequency and quantity must thus be made using partial information, the record of purchases from one retailer.

The second challenge is to account for the heterogeneity of customers. Historically, many researchers have attempted this by incorporating customer level information such as demographics. Even though some online retailers request such information from their customers (household size, income), the data is typically noisy as households are increasingly reluctant to divulge personal information. Furthermore, demographic data often has little informative value for differentiating customers (Rossi,

McCulloch, and Allenby, 1996). An alternative is to estimate models with customer-specific parameters, where household purchase histories provide the information on customer heterogeneity. The online retailer needs therefore a joint model of purchase frequency and purchase quantity, where household-specific parameters can be estimated using purchase information from a single retailer.

We present in this article a model that is specially suited for such an online environment and we illustrate it by modeling the household's joint decision of *when* to shop and *how much* to spend at the store. Our approach is to consider separately 1) the distribution of quantity conditional on purchase timing and 2) the marginal distribution of purchase timing. Though the model is presented as a tool to predict the joint quantity and timing decision of the household, it is flexible enough to be applied to a variety of decision problems (see Borle et al, 2002 for one such application).

We use an integer-valued probability distribution, the Conway-Maxwell-Poisson (henceforth called the CM-Poisson, Conway and Maxwell 1961), as we develop a model of the joint distribution of purchase quantity and purchase frequency. This distribution is a generalization of the Poisson and introduces a second parameter, ν , in addition to the usual Poisson mean λ . The distribution nests the usual Poisson ($\nu = 1$), the geometric ($\nu = 0$), and the Bernoulli ($\nu = \infty$). Values of ν between 0 and 1 permit modeling of over-dispersion, a frequent issue in using the Poisson (Breslow, 1990; Dean, 1992). In this paper, the CM-Poisson is used to model the week of next purchase.

Section 2 describes our data and provides descriptive statistics. In Section 3 we present our hierarchical model (the estimation is discussed in the appendix). Section 4 contains the results of the model, comparing it with various alternative models. We present concluding remarks in Section 5.

2. THE DATA

The data, provided by an online grocery/delivery service, contain all purchases by customers who joined the service from January 1996 through October 1998 in a large metropolitan area. The data includes a census of 14,167 customers, along with their complete purchase histories. We select a random sample of 529 households for the purpose of model estimation, producing a total of 6145 inter-delivery

times and 6674 purchase quantities. A random sample of 500 households was also selected for predictive testing of the models.

Insert Figure 1 here

Figure 1 displays a histogram plot of the inter-delivery times for the 529 households in the estimation sample. Over 90% of these inter-delivery times are less than or equal to 42 days (6 weeks), and about 2.4% lie beyond 100 days. Thus we see that the empirical distribution of the inter-delivery times has most of its mass in the initial region of the distribution. Nevertheless the tail region is not insignificant. The modes correspond to inter-delivery times that are multiples of 7 days, indicating that there is a strong tendency for customers to request deliveries in intervals of a week (or multiples of a week). In part, the weekly tendency could correspond to the planning and purchase schedule of individual households, which may plan meals and household inventory on a weekly basis.

Insert Figure 2 here

Figure 2 displays a scatter plot of the average inter-delivery time for each of the 529 customers versus average purchase quantity. In addition to the strong weekly component of the data, figures 1 & 2 also illustrate the great heterogeneity of the customers. While a large portion of the inter-delivery times are less than or equal to 1 week (27.8%), an equally large portion are greater than 3 weeks (26.4%). Also, the average purchase quantities of the households range from \$ 9.80 to \$ 497.90. The pictured heterogeneity is a result of variation in household loyalty to the online grocer (the share of their total grocery purchases that they make at that retailer), variation in household sizes, and variation in the household's purchase habits regarding the balance of frequency and quantity. The online grocer ideally desires to increase loyalty (share of household) without necessarily increasing purchase frequency. Because the online grocer has access to individual-specific data, the retailer ideally can design individual-specific promotions that encourage households to purchase larger quantities. Such tailored promotions are likely to require household-specific model parameters.

In summary, a simple analysis of the data reveals the need for a joint model of purchase frequency and quantity that allows for cross-sectional heterogeneity and the weekly component of the data. In addition, the model must allow for inferences about the behavior of individual customers and account for the long and significant tail observed in the empirical distribution. In the next section, we introduce a model that addresses each of these issues.

3. THE MODEL

We observe the inter-delivery time t_{ij} in days as well as the corresponding purchase quantity Q_{ij} in dollars, where i and j index a household and its observation respectively. In order to account for the strong week effect, we define w_{ij} and d_{ij} , the week and day components of inter-delivery time such that,

$$t_{ij} = 7w_{ij} + d_{ij}, \quad (1)$$

where $w_{ij} = 0, 1, \dots$ measures the inter-delivery time in weeks rounded off to the nearest week, and d_{ij} , which takes values from the set $\{-3, -2, \dots, 2, 3\}$, is the days around $w_{ij} > 0$. For the scant observations where $w_{ij} = 0$, d_{ij} takes values from the set $\{1, 2, 3\}$. Due to the relatively low probability for $w_{ij} = 0$, we model that event as a special case, using a simple mixture distribution. For $w_{ij} > 0$, we assume w_{ij} to be distributed CM-Poisson (CMP); thus,

$$w_{ij} \sim \begin{cases} (1 - \pi_{ij}) \text{CMP}_{w_{ij}}(\lambda_{ij}, \nu) & \text{for } w_{ij} > 0 \\ \pi_{ij} & \text{for } w_{ij} = 0 \end{cases}, \quad (2)$$

where (λ_{ij}, ν) are the parameters of the CM-Poisson distribution (explained below) and π_{ij} is the probability that $w_{ij} = 0$. We estimate the various parameters of our model using four MCMC samplers (Casella and George, 1992; Gelfand and Smith, 1990). The appendix lays out the full conditional distributions associated with these samplers and Figure 3 is a schematic representation of these four samplers. In the following subsections we lay out the models associated with each.

Insert Figure 3 here

3.1. Sampler 1: Inter-delivery Weeks 1 and Greater

The probability mass function of the CM-Poisson is as follows,

$$P_z(\lambda, \nu) = \left[\sum_{k=0}^{\infty} \frac{\lambda^k}{(k!)^\nu} \right]^{-1} \frac{\lambda^z}{(z!)^\nu}, \quad z = 0, 1, \dots \quad (3)$$

(we model $z_{ij} = (w_{ij} - 1)$ so that the domain of z_{ij} is non-negative integers).

The CM-Poisson is a distribution with thicker or thinner tails than the Poisson and nests the Poisson as a special case (with $\nu=1$). It is defined over positive integers and is flexible in representing a variety of shapes. Though its moments do not have closed form expressions, modern day computing overcomes this inconvenience, allowing us to utilize the advantages that this distribution entails over simpler discrete distributions, especially for thick tailed empirical data. The parameter λ , given ν , is proportional to the expected inter-delivery time, making it a measure of central tendency. In our model, we incorporate inter-temporal as well as cross sectional (across household) variation in λ .

Temporal variation in the expected inter-delivery time for a particular household is modeled by relating the central tendency parameter λ_{ij} to covariates via a multiplicative model,

$$\lambda_{ij} = \lambda_i \delta_1^{x_{1ij}} \delta_2^{x_{2ij}} \dots \delta_k^{x_{kij}} \quad (4)$$

where $x_{1ij}, x_{2ij} \dots x_{kij}$ are time-varying covariates measured in logarithmic form. This parameterization was proposed by Allenby, Leone, and Jen (1999) in the context of the generalized gamma distribution; it has the advantage of easy interpretation as well as estimation. We use the logarithm of one lag of the purchase quantity as the time varying covariate. Thus, a value $\delta = 1$ implies that one lag of the purchase quantity has no effect on the inter-delivery time. For $\delta > 1$, lagged quantity has a positive effect (increase in inter-delivery time); for $\delta < 1$, it has a negative effect (decrease in inter-delivery time). We incorporate household heterogeneity in the expected inter-delivery time by allowing

$$\lambda_i \sim \text{gamma}(\alpha, \beta), \quad (5)$$

and we specify gamma priors on δ , ν , α and an inverse gamma prior on β .

3.2. Sampler 2: Inter-delivery Week 0

While the first sampler focused on a model of deliveries that occur at least 4 days after the

previous delivery, the second sampler estimates the model of deliveries that occur within 3 days of the previous delivery, those cases where $w_{ij}=0$. The probability that $w_{ij} = 0$, π_{ij} , is estimated using the following logit transformation,

$$\pi_{ij} = \frac{1}{1 + \exp(A_i + Bx_{ij})}, \quad (6)$$

where A_i, B are parameters and x_{ij} is a covariate. Heterogeneity across households is modeled by allowing

$$A_i \sim \text{normal}(\psi, \theta), \quad (7)$$

where ψ and θ are mean and variance parameters. The time varying covariate x_{ij} in our model is the logarithm of one lag of the purchase quantity. The coefficient B can be interpreted as the effect of the covariate (one lag of the purchase quantity) on π_{ij} . For values of B greater than 0 a higher purchase quantity would imply a lower probability of the household ordering a next delivery within the next 3 days (i.e. $w_{ij} = 0$). Values less than 0 would have an opposite effect, and a value equal to 0 would imply that the purchase quantity has no bearing on the probability of ordering the next delivery within 3 days. We specify a normal prior on B and conjugate priors on ψ and θ (normal and inverse gamma respectively).

3.3. Sampler 3: Inter-delivery Days

To complete our specification of the inter-delivery time model, we model a 7-dimension vector \mathbf{d}_{ij} (3 dimensions in case when $w_{ij} = 0$), where \mathbf{d}_{ij} is a binary mapping of d_{ij} (the number of days around w_{ij}).

The following logit transformation is used,

$$p(d_{ij}^k = 1) = \frac{\exp(D_k^0)}{\sum_{k=1}^3 \exp(D_k^0)}, \quad D_3^0 = 0, \quad \text{for } w_{ij} = 0 \quad (8)$$

$$p(d_{ij}^k = 1) = \frac{\exp(D_k^1 + \phi_k l_{ij})}{\sum_{k=1}^7 \exp(D_k^1 + \phi_k l_{ij})}, \quad D_4^1 = \phi_4 = 0, \quad \text{for } w_{ij} \geq 1 \quad (9)$$

where k indexes the day of the week (1,2,...,7) and l_{ij} is the ‘‘lateness’’ variable defined as follows,

$$\begin{aligned}
l_{ij} &= \frac{\text{modulo}\{w_{ij}-E^*(w_{ij}), E^*(w_{ij})\}}{\max\left\{1, \text{floor}\left(\frac{w_{ij}-E^*(w_{ij})}{E^*(w_{ij})}\right)\right\}}, & \text{when } w_{ij}-E^*(w_{ij}) > 0 \\
l_{ij} &= 0 & \text{when } w_{ij}-E^*(w_{ij}) = 0 \\
l_{ij} &= \{w_{ij}-E^*(w_{ij})\} \frac{w_{ij}}{E^*(w_{ij})} & \text{when } w_{ij}-E^*(w_{ij}) < 0
\end{aligned} \tag{10}$$

$E^*(w_{ij})$ is the expected value of w_{ij} (from *Sampler 1*) rounded off to the closest integer. The modulo function is defined as $\text{modulo}(a, b) = a - b\text{floor}(a/b)$.

The “lateness” variable (l_{ij}) builds in a dependency of the ‘day’ component of the inter-delivery time with the ‘week’ component (a dependency between *Sampler 1* and *Sampler 3*). l_{ij} is > 0 whenever the household is “late” in the inter-delivery time, i.e. the actual time (the week component) is greater than the expected week component. Similarly a negative value of l_{ij} indicates that the household is “earlier” than the expected inter-delivery time (the week component). The specific functional form of l_{ij} (equation 10) serves one more purpose viz.; it minimizes the effect of outliers in the data. This is accomplished by tying down the difference $[w_{ij} - E^*(w_{ij})]$ using the modulo function and utilizing the “weekly” periodicity observed in the empirical distribution.

The parameters D_k^I and ϕ_k are intercept and slope parameters, and we specify normal priors on them (we also specify normal priors on D_k^O). The parameter ϕ_k can be interpreted as the effect of ‘week’ component on the ‘day’ component of the inter-delivery time. Higher values of ϕ_1 , ϕ_2 and ϕ_3 as compared to ϕ_5 , ϕ_6 and ϕ_7 would imply that if the ‘week’ purchase cycle gets delayed (i.e. l_{ij} is positive) the household would be more likely to shift the delivery earlier in the week.

Thus, given w_{ij} , every household has different probabilities for the ‘days’ component of its inter-delivery time. Such a specification allows the model to predict inter-delivery times in units of days and may particularly be relevant for online grocery/delivery firms that need to plan deliveries on a daily basis.

3.4. Sampler 4: Purchase Quantity

We model purchase quantity conditional on inter-delivery time. If the grocery purchase is an inventory replenishment, an especially long inter-delivery time would likely be followed by a larger order quantity. We assume that,

$$\log(Q_{ij}) \sim \text{Normal}(\boldsymbol{\gamma}_i \mathbf{t}_{ij}, \sigma^2), \quad (11)$$

Q_{ij} is the purchase quantity in dollars, $\boldsymbol{\gamma}_i = (\gamma_{0i}, \gamma_{1i})$, $\mathbf{t}_{ij} = (1, \log t_{ij})'$, t_{ij} being the inter-delivery time and σ^2 is the variance parameter.

This parameterization has a pleasing managerial interpretation. For the customer who orders groceries exclusively via the delivery service, we would expect the purchase quantity to be roughly proportional to the time since the previous purchase, i.e. $\gamma_1 \approx 1$. For the customer who chooses a supplier at random, where the online service is one of multiple suppliers, the dependency between purchase quantity and inter-delivery time would be almost nonexistent, $\gamma_1 \approx 0$. In marketing terms, this parameter maps out a space for the loyalty of the customer.

Heterogeneity across households is modeled by letting

$$\boldsymbol{\gamma}_i \sim \text{Normal}(\bar{\boldsymbol{\gamma}}, \mathbf{V}_\gamma) \quad (12)$$

We complete the Bayesian specification by specifying conjugate priors on σ^2 , $\bar{\boldsymbol{\gamma}}$ and \mathbf{V}_γ .

3.5. Prior Specifications

The appendix lists the MCMC estimation algorithm as well as the priors used in the estimation. For sampler 1, we need to specify priors on the ν , α , β and δ parameters. ν is the decay parameter and is defined over the positive real line, *a priori*, there is little information on the range of values ν can take; however, very high values of ν seem unlikely given the dispersion observed in our data. The prior distribution on ν , a gamma(2,1) with a mode at 1 is a reasonable representation of our belief on the values ν can take. The parameters α and β determine the cross-sectional heterogeneity distribution of λ_i , which, in turn along with the coefficient for the lagged purchase quantity (δ), determines λ_{ij} , the central tendency parameter of inter delivery time. Again, *a priori* there is little information to form ‘sharp’ priors on λ_i and

we specify priors on α and β that can accommodate a wide range of values for λ_i . The parameter δ is the effect of the logarithm of one lag of purchase quantity on inter-delivery time and it can take only positive values. We expect an increase in purchase quantity to delay the subsequent purchase (i.e. $\delta > 1$); however, we do not have any specific expectations on its magnitude except that very large values seem unlikely. The prior distribution on δ , the gamma(5,5), can accommodate a wide range of values for δ .

For sampler 2, priors are needed for the ψ , θ and B parameters. The parameters ψ and θ specify the mean and variance respectively of the heterogeneity distribution over A_i . The A_i 's can be interpreted as the 'inherent proclivity' of the households to order deliveries within 3 days ($w_{ij} = 0$) of their previous delivery. The A_i 's along with B (equation 6) determine the 'overall' variation in π_{ij} (which is the probability that $w_{ij} = 0$), the parameter B being the effect of the logarithm of one lag of purchase quantity on π_{ij} . We expect a majority of households to have a low inherent proclivity to order within 3 days and the parameter B to be positive, though we do not have sharp expectations about its magnitude. The prior distributions specified for ψ , θ and B represent these prior expectations.

The parameters D_k^0 , D_k^1 and ϕ_k (sampler 3) specify the distribution of the 'day' component of the inter-delivery time, D_k being the intercept and ϕ_k the slope coefficient in equation (9). Again, *a priori* there is little information on their magnitudes and we utilize the normal(0,10) for these parameters.

For sampler 4, we center our prior distribution about zero on $\bar{\gamma}$, the two-dimensional expectation of γ_i (which maps out the loyalty of the households to the online grocer). The prior specified has a large variance to accommodate a wide range of $\bar{\gamma}$. The parameter \mathbf{V}_γ , controls the shrinkage (heterogeneity) of customer loyalty and σ^2 is the variance of the purchase quantities. The inverse wishart prior on \mathbf{V}_γ and the inverse gamma prior on σ^2 is consistent with our lack of specific prior knowledge on these parameters.

4. ANALYSIS OF ONLINE GROCERY DATA

4.1 Parameter Estimates

Table 1 reports the point estimates for parameters that are not specific to individual households.

Insert Table 1 here

The estimated heterogeneity distribution over the λ_i 's has a mean 0.6955 ($\alpha\beta$) and variance 0.0730 ($\alpha\beta^2$). The λ_i 's combine with the δ and x_{ij} 's to produce the *overall* heterogeneity in the inter-delivery times [equations (2), (3) and (4)], while the λ_i 's themselves can be interpreted as capturing the *inherent* heterogeneity in mean inter-delivery times across households.

The impact of lagged quantity on the expected inter-delivery time (δ) is fairly small from a managerial perspective. An additional 10% of purchase amount on the previous order increases λ_{ij} (which is equal to $\lambda_i \delta^{x_{ij}}$) by a small 0.22% ($\delta^{\log 1.1}$). This translates to an increase in expected inter-delivery times in the range 0.23% to 1.27%, the median being 0.48%. Discussions with managers revealed that almost all of the online customers also purchase from a traditional local retailer. Potentially δ is small in magnitude because households purchase much of their pantry inventory elsewhere, weakening the correlation between inter-delivery time and previous purchase quantity.

The parameter ν is a measure of the decay of purchase probabilities of the weeks. For all ν , the ratio $p(w_{ij}=n+1)/p(w_{ij}=n)$ is equal to λ_{ij}/n^ν . When $\nu=1$, the distribution is the (regular) Poisson. The thick tail of the empirical distribution of inter-delivery times leads to a slower decay (smaller ν) for the CM-Poisson. However, the empirical histogram has a faster decay in the initial 3 to 4 weeks. Given the smaller value of ν , an initial fast decay is achieved by lower λ_{ij} 's. This result illustrates the flexibility of the CM-Poisson in modeling data with such characteristics. It is flexible enough to incorporate long tails and at the same time a relatively faster decay in the initial parts of the distribution.

The average posterior probability across households that $w_{ij}=0$ (i.e. the inter-delivery time is within 3 days) calculated from (6) is extremely close to 0 for most households, indicating that few households purchase within 3 days of a previous purchase. Further, the impact of purchase quantity of the lagged (i.e. previous) purchase on π_{ij} (which is the probability that $w_{ij}=0$) is slight from a managerial perspective. In the span of purchase quantities observed in the dataset, an additional 10% of purchase on

the previous order decreases the expected value of π_{ij} in the range 2.1% to 6.2% [from (6), using the posterior means for A_i and B].

The remaining parameters of the inter-delivery time portion of the model in Table 1 (D_k^0 's, D_k^1 's and ϕ_k 's), characterize the distribution of the day of the week on which the delivery was ordered. The results are as expected, where the largest probabilities occur near days corresponding to weekly inter-delivery times. Further, amongst the ϕ_k 's only ϕ_1 , ϕ_2 and ϕ_3 have 'significant' positive values, thus implying that households tend to pre-poner their deliveries within the week if they get late in their 'weekly' delivery cycles.

As for purchase quantity, the posterior population means for the intercept and slope of the linear log quantity equation (equations 12) are 4.4054 and 0.0821 respectively. The average order size for single-week inter-delivery times is $\exp(4.4054 + 0.0821 \cdot \log 7) = \96.08 , while on average, doubling the inter-delivery time results in $\exp(0.0821 \cdot \log 2) = 1.0586$, or a 5.86% increase in purchase quantity. These results reinforce the views expressed in our discussion with managers that almost all online customers also purchase from traditional local retailers. In marketing terms, these households have little "loyalty" to the online grocer. Perhaps, it is still early days for internet grocers, and the internet as a channel for grocery purchases may yet not have grown firm roots.

4.2 Model Comparisons and Predictive Performance

Along with the proposed model (model 1), we also estimate six other competing models. Two of these are nested models, the Poisson (model 2: $\nu = 1$, corresponding to a regular Poisson distribution for w_{ij}), the Geometric without covariates (model 3: $\nu = 0$, corresponding to a Geometric distribution for w_{ij}). The Geometric distribution restricts λ_{ij} (which is equal to $\lambda_i \delta^{x_{ij}}$, x_{ij} being one lag of the log purchase amount) to values less than 1. In this case, the parameter δ loses much of its interpretive usefulness in terms of effect on inter-delivery times for a given increase in lagged purchase quantity (keeping λ_i fixed). For this reason we estimate this model without the covariates (i.e. keeping $\delta = 1$). For comparison purposes, we also estimate the CM-Poisson model without the covariates (model 4). The fifth model is an

aggregate model with no individual level parameters or covariates (model 5). Finally the last two models (models 6 and 7) are without the day-of-week effect i.e. where the inter-delivery time is not segregated into a week and a day component.

For sake of brevity, we do not report the parameter estimates from these models. However, Table 2 reports the log marginal densities and also compares the predictive performance of these models. The log marginal densities reported are for the marginal model of inter-delivery time (samplers 1, 2 and 3) and have been calculated as suggested by Newton and Raftery (1994). The mean absolute deviation (MAD) statistic (difference between actual and predicted inter-delivery times) is used as a metric to compare predictive accuracy across various models. We consider two prediction scenarios and since all the models have the same ‘quantity’ model we restrict our analysis to the ‘timing’ models.

Insert Table 2 here

Considering the log marginal densities the data favors the CM-Poisson model over all the other models (Kass and Raftery, 1995). Although the Poisson fits the center of the data quite well, it does not fit the extremes as well as the CM-Poisson or the Geometric model. For marketing applications concerning traditional brick-and-mortar retailers, the households in the tail are less interesting because they are the less profitable infrequent customers. By deleting infrequent purchasers, previous studies have been able to use known distributions with thinner tails (Gupta, 1991). Here, where the on-line grocer especially profits from infrequent customers, a distribution allowing added flexibility in the tails becomes important.

Figure 4 illustrates the cumulative hazard functions for the CM-Poisson and the Poisson models. These functions have been computed from the survivor curve which was produced by integrating over the heterogeneity distribution over λ_{ij} . As seen, in the case of CM -Poisson there is a big departure from the Poisson’s hazard, again pointing to the usefulness of the CM-Poisson in modeling long tail empirical distributions.

Insert Figure 4 here

The week/day periodicity observed in the empirical distribution is an important aspect of the data; ignoring it leads to substantial differences in the estimated models. We estimate two such models without the week effect i.e. the inter-delivery time is not broken up into a week component and a day component. Model 6 has t_{ij} , the inter-delivery time in days, modeled as CM-Poisson [$t_{ij} \sim CMP(\lambda_{ij}, \nu)$], where $\lambda_{ij} = \lambda_i \delta^{x_{ij}}$] and Model 7 has it modeled as lognormal. Again, considering the log marginal densities, the data favors the proposed CM-Poisson model (model 1) over these simpler versions.

The lognormal model yields expected inter-delivery times for individual households in the range 4 to 89 days (the mean across all households being 22 days); the middle 95% interval being 8 to 52 days. The actual times as observed in the data range from 3 to 274 days, the 95% interval being 6 to 90 days and the mean across all households being 33 days. In comparison the proposed model (model 1) yields expected inter-delivery times per household in the range 6 to 212 days with the 95% interval being 10 to 83 days and the mean across all households being 29 days. The lognormal model does not capture as much of the variation in the data as does the CM-Poisson model.

We also compare our results from the proposed model (model 1) with that of a model using aggregate data. Since aggregate parameters are easier for a retailer to use, the extra effort for household-specific estimates is warranted only if the conclusions are different. We have included a comparison with one such model, a pooled model without household level parameters (Model 5). In terms of log marginal densities, the data does not favor the pooled model; the model is one of the worst models based on this measure. The expected inter-delivery time is 21 days for this model for all the households as compared to 29 days (across all households) in our proposed model (the data average is 33 days). Further, comparing a simulation of inter-delivery times from both the models with the actual inter-delivery times, 22% of these simulated times were misclassified in the CM-Poisson model (model 1), as opposed to 32% in the pooled model (model 5). The 95% quantile for these simulations were 62 and 55 days respectively for models 1 and 5 (the actual empirical distribution has 69 days as the 95% quantile point). Aggregation results from the household level estimates of the proposed model outperforms the estimates from an aggregate model.

The predictive performance of the models was compared in two scenarios. In the first scenario we re-estimated all the seven models in Table 2 using all but the last observation for each individual household. We kept the last observation for predictive analysis. This resulted in 414 households being used in the estimation (the remaining 115 households had single observations). In the second scenario a random holdout sample of 500 households was used for predictive testing of the models. In the second scenario the predictions are based on the point estimates reported in Table 1., i.e. we do not update the parameters but do use the data to generate individual-level estimates of expected inter-delivery times. This leads to a very fast algorithm and is also how an online retailer is likely to use the model in practice.

In general the predictive error is large across all models. In scenario 1 the CM-Poisson model does not result in very different predictions than the simpler, Poisson or Geometric models. Part of the reason is that the models in some sense “over-fit” the data. Thus, the relative advantage of the CM-Poisson distribution is difficult to discern in predictions within the fitted sample. However, in scenario 2 in which we use a holdout sample, marked improvements are seen in predictions using the CM-Poisson model, there is a striking improvement of the CM-Poisson model relative to the Poisson model (model 2). This gain in prediction is primarily due to the inflexibility of the Poisson to model ‘over-dispersed’ data. The CM-Poisson distribution has the added flexibility to model extreme units in the distribution and thus leads to more accurate predictions.

In summary, the proposed model performs better than the other models in holdout sample predictions; however, the predictive performance is very ‘noisy’ for all the models across both scenarios. A ‘richer’ covariate set (one incorporating various elements of the marketing mix) is likely to increase the predictive accuracy of the models.

5. CONCLUDING REMARKS

The pick, pack and delivery systems are perhaps the largest capital outlay that an online retailer makes, especially so in the case of an online grocer where the presence of perishable products makes it an even more important issue. Unlike the traditional brick-and-mortar grocery store, which can benefit from

sales models developed in the marketing literature, an online store would need a different model. This is primarily because of the different cost structure of the online store where the pick, pack and delivery is an important part of overall costs and thus would require a model that can segregate the overall sales into frequency and quantity dimensions, a model that jointly models the consumer decisions of *when* and *how much* to buy. An online store can gainfully employ such a model not only to target individual households but also to plan for its logistics.

In this article we develop a model for the joint distribution of the household's timing and quantity decisions and apply it to data from an online grocery store. Our approach is to model the marginal distribution of purchase timing (inter-delivery times) and the distribution of purchase quantity conditional on purchase timing. The model is estimated using purchase histories of 529 households from an online grocery delivery company. We propose a hierarchical Bayes model that disentangles the weekly and daily components of inter-delivery times and models them separately, thereby accounting for strong periodicity observed in the data. We use the distribution CM-Poisson, which we find useful to fit data in the tail regions (extremely frequent and infrequent purchasers), a feature that is especially important for an online grocery/delivery company whose costs are directly tied to purchase frequency. Use of the CM-Poisson increased predictive accuracy of the model, as simpler models did not adequately account for consumer heterogeneity.

The proposed model has the advantage that it helps predict the inter-delivery times at a disaggregate level (the unit being one day), a feature that can be used by the online retailer to optimize its delivery schedules on a daily basis. The 'day' component of inter-delivery time is dependent on the 'week' component and our results show that households who order 'late' on their 'week' component (from the expected value of the 'week' component) tend to make-up for it in the 'day' component of the inter-delivery time.

One of the drawbacks of the results that we obtain is that the covariates are not strongly predictive. What is encouraging about the covariates is that their algebraic signs are as expected. However, the model specification is flexible enough to incorporate further covariates. The model

developed provides a useful approach to build models of joint distribution of frequency and purchase amounts. Though this paper illustrates the model in one setting, many other opportunities exist for this type of joint distribution model.

APPENDIX

The estimation proceeds by running four MCMC samplers (Casella and George, 1992; Gelfand and Smith, 1990). Figure 3 gives a schematic view of these samplers.

Sampler 1	Sampler 2	Parameter	Priors
a) $\Pi[\lambda_i / \{w_{ij}\}, \{x_{ij}\}, \{I_{ij}\}, \delta, \nu, \alpha, \beta]$	a) $\Pi[A_i / \{x_{ij}\}, \{I_{ij}\}, B, \psi, \theta]$	ν	Gamma(2,1)
b) $\Pi[\delta / \{w_{ij}\}, \{x_{ij}\}, \{I_{ij}\}, \{\lambda_i\}, \nu, a_1, b_1]$	b) $\Pi[B / \{x_{ij}\}, \{I_{ij}\}, A_i, c_1, d_1]$	α	Gamma(5,5)
c) $\Pi[\nu / \{w_{ij}\}, \{x_{ij}\}, \{I_{ij}\}, \{\lambda_i\}, \delta, a_2, b_2]$	c) $\Pi[\psi / \{A_i\}, \theta, c_2, d_2]$	β	Inv gamma(2.5,0.5)
d) $\Pi[\alpha / \{\lambda_i\}, \beta, a_3, b_3]$	d) $\Pi[\theta / \{A_i\}, \psi, c_3, d_3]$	δ	Gamma(5,5)
e) $\Pi[\beta / \{\lambda_i\}, \alpha, a_4, b_4]$		ψ	Normal(0,10)
		θ	Inv gamma(5,5)
		B	Normal(0,10)
		D_1^0, D_2^0, D_3^0	Normal(0,10)
		$D_1^1, D_2^1, \dots, D_7^1$	Normal(0,10)
		$\phi_1, \phi_2, \dots, \phi_7$	Normal(0,10)
		$\bar{\gamma}$	MVN([0,0], 10I)
		\mathbf{V}_γ	Inv wishart(1,15I,10)
		σ^2	Inv gamma(2.5,0.5)
Sampler 3	Sampler 4		
a) $\Pi[D_k^0 / \{\mathbf{d}_{ij}\}, \{I_{ij}\}, \{l_{ij}\}, e, f]$	a) $\Pi[\gamma_i / \{\log Q_{ij}\}, \{\hat{\mathbf{t}}_{ij}\}, \sigma^2, \bar{\gamma}, \mathbf{V}_\gamma]$		
b) $\Pi[D_k^1 / \{\mathbf{d}_{ij}\}, \{I_{ij}\}, \{l_{ij}\}, \{\phi_k\}, e, f]$	b) $\Pi[\sigma^2 / \{\log Q_{ij}\}, \{\hat{\mathbf{t}}_{ij}\}, \{\gamma_i\}, m, n]$		
c) $\Pi[\phi_k / \{\mathbf{d}_{ij}\}, \{I_{ij}\}, \{l_{ij}\}, \{D_k^1\}, e, f]$	c) $\Pi[\bar{\gamma} / \{\gamma_i\}, \mathbf{V}_\gamma, \gamma_0, \mathbf{V}_0]$		
	d) $\Pi[\mathbf{V}_\gamma / \{\gamma_i\}, \bar{\gamma}, k, \mathbf{G}, g]$		

The first sampler corresponds to (3), (4) and (5) modeling the ‘week’ part of the inter-delivery time. The second sampler corresponds to (6) and (7) modeling the probability that $w_{ij} = 0$. The third sampler corresponds to (8) to (10) modeling the ‘day’ part of the inter-delivery time. Finally, the fourth sampler corresponds to (11) and (12) modeling the purchase quantity. Figure 1 gives a schematic view of these four samplers.

Sampler 1 proceeds by recursively drawing from the following set of full conditional distributions.

$$\Pi[\lambda_i / \{w_{ij}\}, \{x_{ij}\}, \{I_{ij}\}, \delta, \nu, \alpha, \beta] \propto \left[\prod_{j=1}^{J_i} \left\{ \sum_{k=0}^{\infty} \frac{(\lambda_i \delta^{x_{ij}})^k}{(k!)^\nu} \right\}^{-I_{ij}} \right] \lambda_i^{\nu \sum_{j=1}^{J_i} (w_{ij}-1) I_{ij}} \lambda_i^{\alpha-1} e^{-\frac{\lambda_i}{\beta}} \quad (13)$$

$$\Pi[\delta / \{w_{ij}\}, \{x_{ij}\}, \{I_{ij}\}, \{\lambda_i\}, \nu, a_1, b_1] \propto \left[\prod_{i=1}^N \prod_{j=1}^{J_i} \left\{ \sum_{k=0}^{\infty} \frac{(\lambda_i \delta^{x_{ij}})^k}{(k!)^\nu} \right\}^{-I_{ij}} \right] \delta^{\nu \sum_{i=1}^N \sum_{j=1}^{J_i} x_{ij} (w_{ij}-1) I_{ij}} \delta^{a_1-1} e^{-\frac{\delta}{b_1}} \quad (14)$$

$$\Pi[\nu / \{w_{ij}\}, \{x_{ij}\}, \{I_{ij}\}, \{\lambda_i\}, \delta, a_2, b_2] \propto \left[\prod_{j=1}^{J_i} \left\{ \sum_{k=0}^{\infty} \frac{(\lambda_i \delta^{x_{ij}})^k}{(k!)^\nu} \right\}^{-I_{ij}} \right] \prod_{j=1}^{J_i} \frac{(\lambda_i \delta^{x_{ij}})^{(w_{ij}-1) I_{ij} \nu}}{\{(w_{ij}-1)!\}^{I_{ij} \nu}} \psi^{a_2-1} e^{-\frac{\nu}{b_2}} \quad (15)$$

$$\Pi[\alpha / \{\lambda_i\}, \beta, a_3, b_3] \propto \frac{\prod_{i=1}^N \lambda_i^{\alpha-1}}{[\Gamma(\alpha)]^N \beta^{N\alpha}} \cdot \alpha^{a_3-1} e^{-\frac{\alpha}{b_3}} \quad (16)$$

$$\Pi[\beta / \{\lambda_i\}, \alpha, a_4, b_4] \sim \text{inv gamma} \left[a_4 + N\alpha, \left(\sum_{i=1}^N \lambda_i + b_4^{-1} \right)^{-1} \right] \quad (17)$$

I_{ij} is an index variable such that $I_{ij} = 1$ when $w_{ij} > 0$, else $I_{ij} = 0$. N is the total number of households, J_i is the total number of observations for the i^{th} household. The parameters (a_1, b_1) , (a_2, b_2) and (a_3, b_3) specify the gamma prior distribution on δ , ν and α respectively, while (a_4, b_4) the inverse gamma prior distribution on β .

The infinite sum series in the distributions is a convergent series and needs numerical approximation.

Sampler 2 proceeds by recursively drawing from the following set of full conditional distributions.

$$\Pi[A_i / \{x_{ij}\}, \{I_{ij}\}, B, \psi, \theta] \propto \left[\prod_{j=1}^{J_i} (1 + \exp(A_i + Bx_{ij}))^{-1} \right] \exp\left(A_i \left[\left(\sum_{j=1}^{J_i} I_{ij} \right) - \frac{.5A_i}{\theta} + \frac{\psi}{\theta} \right] \right) \quad (18)$$

$$\Pi[B / \{x_{ij}\}, \{I_{ij}\}, A_i, c_1, d_1] \propto \left[\prod_{n=1}^N \prod_{j=1}^{J_i} (1 + \exp(A_i + Bx_{ij}))^{-1} \right] \exp\left(\left\{ B \sum_{i=1}^N \sum_{j=1}^{J_i} x_{ij} I_{ij} \right\} - \frac{(B - c_1)^2}{2d_1} \right) \quad (19)$$

$$\Pi[\psi / \{A_i\}, \theta, c_2, d_2] \sim \text{normal} \left[\frac{c_2\theta + d_2 \sum_{i=1}^N A_i}{\theta + Nd_2}, \left(\frac{1}{d_2} + \frac{N}{\theta} \right)^{-1} \right] \quad (20)$$

$$\Pi[\theta / \{A_i\}, \psi, c_3, d_3] \sim \text{invgamma} \left[\frac{N}{2} + c_3, \left\{ .5 \sum_{i=1}^N (A_i - \psi)^2 + \frac{1}{d_3} \right\}^{-1} \right] \quad (21)$$

(c_1, d_1) are the parameters of the normal prior distribution on B . (c_2, d_2) and (c_3, d_3) are the parameters of the conjugate normal and inverse gamma prior distributions on ψ and θ respectively.

Sampler 3 draws recursively from the following distributions.

$$\Pi[D_k^0 / \{\mathbf{d}_{ij}\}, \{I_{ij}\}, \{l_{ij}\}] \propto \left[\prod_{i=1}^N \prod_{j=1}^{J_i} \left\{ \frac{\exp(\mathbf{d}_{ij}^k D_k^0)}{\sum_{k=1}^3 \exp(D_k^0)} \right\}^{1-I_{ij}} \right] \exp\left(-\frac{D_k^0 - e}{2f}\right) \quad (22)$$

$$\Pi[D_k^1 / \{\mathbf{d}_{ij}\}, \{I_{ij}\}, \{l_{ij}\}, \{\phi_k\}] \propto \left[\prod_{i=1}^N \prod_{j=1}^{J_i} \left\{ \frac{\exp(\mathbf{d}_{ij}^k D_k^1)}{\sum_{k=1}^7 \exp(D_k^1 + \phi_k l_{ij})} \right\}^{I_{ij}} \right] \exp\left(-\frac{D_k^1 - e}{2f}\right) \quad (23)$$

$$\Pi[\phi_k / \{\mathbf{d}_{ij}\}, \{I_{ij}\}, \{l_{ij}\}, \{D_k^1\}] \propto \left[\prod_{i=1}^N \prod_{j=1}^{J_i} \left\{ \frac{\exp(\mathbf{d}_{ij}^k \phi_k l_{ij})}{\sum_{k=1}^7 \exp(D_k^1 + \phi_k l_{ij})} \right\}^{I_{ij}} \right] \exp\left(-\frac{\phi_k - e}{2f}\right) \quad (24)$$

(e,f) are the parameters of the normal prior distributions on the D_k^0 , D_k^1 and ϕ_k parameters.

Sampler 4 draws recursively from the following distributions.

$$\begin{aligned} \Pi[\boldsymbol{\gamma}_i / \{\log Q_{ij}, \hat{\mathbf{t}}_{ij}, \sigma^2, \bar{\boldsymbol{\gamma}}, \mathbf{V}_\gamma\}] \propto \\ \exp\left(-\frac{1}{2}(\boldsymbol{\gamma}_i - \bar{\boldsymbol{\gamma}})\mathbf{V}_\gamma^{-1}(\boldsymbol{\gamma}_i - \bar{\boldsymbol{\gamma}})' - \frac{1}{2\sigma^2} \sum_{j=1}^{J_i} (\log Q_{ij} - \boldsymbol{\gamma}_i \hat{\mathbf{t}}_{ij})^2\right) \end{aligned} \quad (25)$$

$$\begin{aligned} \Pi[\sigma^2 / \{\log Q_{ij}, \hat{\mathbf{t}}_{ij}, \{\boldsymbol{\gamma}_i\}, m, n\}] \sim \\ \text{inv gamma}\left[\frac{1}{2} \sum_{i=1}^N J_i + m, \left\{\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^{J_i} (Q_{ij} - \boldsymbol{\gamma}_i \hat{\mathbf{t}}_{ij})^2 - n^{-1}\right\}^{-1}\right] \end{aligned} \quad (26)$$

$$\Pi[\bar{\boldsymbol{\gamma}} / \{\boldsymbol{\gamma}_i, \mathbf{V}_\gamma, \boldsymbol{\gamma}_0, \mathbf{V}_0\}] \sim \text{Normal}[\boldsymbol{\gamma}_n, \mathbf{V}_n] \quad (27)$$

where

$$\boldsymbol{\gamma}_n = \frac{\boldsymbol{\gamma}_0 \mathbf{V}_0^{-1} + \sum_{i=1}^N \boldsymbol{\gamma}_i \mathbf{V}_\gamma^{-1}}{\mathbf{V}_0^{-1} + N\mathbf{V}_\gamma^{-1}}, \quad \mathbf{V}_n = \mathbf{V}_0^{-1} + N\mathbf{V}_\gamma^{-1}$$

$$\begin{aligned} \Pi[\mathbf{V}_\gamma / \{\boldsymbol{\gamma}_i, \bar{\boldsymbol{\gamma}}, k, \mathbf{G}, \mathbf{g}\}] \sim \\ \text{Inv wishart}\left[\sum_{i=1}^N (\boldsymbol{\gamma}_i - \bar{\boldsymbol{\gamma}})'(\boldsymbol{\gamma}_i - \bar{\boldsymbol{\gamma}}) + \mathbf{G} + \frac{kN}{N+k}(\bar{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0)'(\bar{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0), N + \mathbf{g}\right] \end{aligned} \quad (28)$$

The parameters (m,n) are that of the inverse gamma prior on σ^2 , $(\boldsymbol{\gamma}_0, \mathbf{V}_0)$ are the parameters of the multivariate normal prior distribution on $\bar{\boldsymbol{\gamma}}$ and $(k, \mathbf{G}, \mathbf{g})$ are the parameters of the inverse wishart prior distribution on \mathbf{V}_γ .

Note that many of the conditional distributions in the above four samplers are from non-conjugate distributions. We employ a random walk Metropolis Hastings algorithm to obtain draws from these distributions, using a truncated normal proposal density with its variance used as a tuning constant (Chib and Greenberg 1995).

REFERENCES

- Allenby, G. M., Arora, N. and Ginter, J. L. (1998), "On the Heterogeneity of Demand", *Journal of Marketing Research*, 35, 384-389.
- Allenby, G. M., Leone, R. P. and Jen, L. (1999), "A Dynamic Model of Purchase Timing With Application to Direct Marketing", *Journal of the American Statistical Association*, 94, 365-374
- Bell, D. R. and Lattin, J. M. (1998), "Shopping Behavior and Consumer Preference for Store Price Format: Why 'Large Basket' Shoppers prefer EDLP", *Marketing Science*, 17, 66-88.
- Boatwright, P. and Nunes, J. C. (2001), "Reducing Assortment: An Attribute-Based Approach", *Journal of Marketing*, 65, 50-63
- Boatwright, P., McCulloch R. and Rossi, P. E. (1999), "Account-Level Modeling for Trade Promotion: An Application of a Constrained Parameter Hierarchical Model", *Journal of the American Statistical Association*, 94, 1063-1073
- Borle, S., Boatwright, P., Kadane, J. B., Nunes, J. C. and Shmueli, G. (2002), "Effect of Product Assortment Changes on Customer Retention", *GSIA Working Paper*, Carnegie Mellon University.
- Breslow, N. (1990) "Tests of Hypotheses in Overdispersed Poisson Regression and Other Quasi-Likelihood Models", *Journal of the American Statistical Association*, 85, 565-571.
- Bucklin, R. E. and Gupta, S. (1999), "Use of UPC Scanner Data: Industry and Academic Perspectives", *Marketing Science*, 18, 247-273.
- Bucklin, R. E. and Lattin, J. M. (1991), "A Two State Model of Purchase Incidence and Brand Choice", *Marketing Science*, 10, 24-39.
- Carlin, B.P. and Chib, S. (1995), "Bayesian Model Choice via Markov Chain Monte Carlo Methods", *Journal of the Royal Statistical Society*, B, 57, 473-484.

- Casella, G. and George, E. I. (1992), "Explaining the Gibbs Sampler", *The American Statistician*, 49, No. 4, 327-335.
- Chib, S. and Greenberg, E. (1995), "Understanding the Metropolis-Hastings Algorithm", *The American Statistician*, Vol. 49, No. 4, 327-335
- Chintagunta, P.K. (1993), "Investigating Purchase Incidence, Brand Choice and Purchase Quantity Decisions of Households", *Marketing Science*, 12, 184-208.
- Conway, R.W. and Maxwell, W.L.. (1961), "A Queueing Model with State Dependent Service Rate", *The Journal of Industrial Engineering*, 12, 132-136.
- Dean, C.B. (1992), "Testing for Overdispersion in Poisson and Binomial Regression Models", *Journal of the American Statistical Association*, 87, 451-457.
- Gelfand, A. E. and Smith, A. F. M. (1990), "Sampling Based Approaches to Calculating Marginal Densities", *Journal of the American Statistical Association*, 85, 398-409
- Gupta, S. (1988), "Impact of Sales Promotions on When, What and How Much to Buy", *Journal of Marketing research*, 25, 342-356.
- Gupta, S. (1991), "Stochastic Models of Interpurchase Time With Time-Dependent Covariates", *Journal of Marketing Research*, 28, 1-15.
- Gupta, S. and Morrison D. G. (1991), "Estimating Heterogeneity in Consumers' Purchase Rates", *Marketing Science*, 10, 264-269.
- Heidelberger, P. and Welch, P. (1983), "Simulation run length control in the presence of an initial transient", *Operations Research*, 31, 1109-1144.
- Jain, D. and Vilcassim, N. J. (1991), "Investigating Household Purchase Timing Decisions: A Conditional Hazard Function Approach", *Marketing Science*, 10, 1-23.

- Jedidi, K., Mela, C. F. and Gupta, S. (1999), "Managing Advertising and Promotion for Long-run Profitability," *Marketing Science*, 18, 1-22.
- Kass, R.E and Raftery, A.E. (1995), "Bayes Factors", *Journal of the American Statistical Association*, 90, 773-795
- Meghir, C. and Robin, J. –M. (1992), "Frequency of purchase and the estimation of demand systems", *Journal of Econometrics*, 53, 53-85.
- Neslin, S. A., Henderson C. and Quelch, J. (1985), "Consumer Promotions and Acceleration of Product Purchases", *Marketing Science*, 4, 147-165.
- Newton, M.A. and Raftery, A.E. (1994), "Approximate Bayesian Inference With the Weighted Likelihood Bootstrap", *Journal of the Royal Statistical Society, Ser. B*, 56, 3-48
- O'Neill, S. and Chu, J. (2001), "Modeling to scale: The Evolution of Online Grocery Retailing", *IBM Global Services eStrategy Report*, July.
- Rossi, P. E., McCulloch, R. E. and Allenby, G. M. (1996), "The Value of Purchase History Data in Target Marketing", *Marketing Science*, 15, 147-165.
- Schmittlein, D. C., Morrison, D. G. and Colombo, R. (1987), "Counting Your Customers: Who Are They and What Will They Do Next?", *Management Science*, 33, 1-24
- Van Praag, B. M. S. and Vermeulen, E. M. (1993), "A Count-Amount Model with Endogenous Recording of Observations", *Journal of Applied Econometrics*, 8, 383-395.

Table 1. Parameter Estimates

Parameters	The CM-Poisson Model
α	6.3064 _{0.5628}
β	0.1049 _{0.0103}
δ	1.0229 _{0.0108}
ν	0.0631 _{0.0086}
ψ	2.5514 _{0.8055}
θ	2.1858 _{0.6869}
B	0.6691 _{0.1744}
$\begin{bmatrix} D_1^0 \\ D_2^0 \\ D_3^0 \end{bmatrix}$	$\begin{bmatrix} -1.3226 & 0.2503 \\ -0.7007 & 0.2087 \\ 0.0000 & \end{bmatrix}$
$\begin{bmatrix} D_1^1 \\ D_2^1 \\ D_3^1 \\ D_4^1 \\ D_5^1 \\ D_6^1 \\ D_7^1 \end{bmatrix}$	$\begin{bmatrix} -1.2628 & 0.0502 \\ -1.0642 & 0.0467 \\ -0.7470 & 0.0418 \\ 0.0000 & \\ -0.7278 & 0.0414 \\ -1.0441 & 0.0489 \\ -1.3043 & 0.0514 \end{bmatrix}$
$\begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \phi_4 \\ \phi_5 \\ \phi_6 \\ \phi_7 \end{bmatrix}$	$\begin{bmatrix} 0.1169 & 0.0414 \\ 0.0781 & 0.0477 \\ 0.0725 & 0.0436 \\ 0.0000 & \\ -0.0111 & 0.0453 \\ -0.0104 & 0.0515 \\ 0.0098 & 0.0557 \end{bmatrix}$
σ^2	0.0955 _{0.0021}
$\bar{\gamma}$	$\begin{bmatrix} 4.4054 & 0.0420 \\ 0.0821 & 0.0157 \end{bmatrix}$

NOTE: Subscripted numbers are posterior standard deviations

Convergence was confirmed via the Bayesian Output Analysis Program (BOA) Version 0.5.0 (Heidelberger and Welch Convergence Diagnostics; Heidelberger and Welch (1983))

For sake of brevity some of the parameters have not been reported.

Table 2. Comparison of Model Performance

	LMD ^a	MAD ^b (scenario 1)	MAD (scenario 2)
<i>Model 1</i>	-21414.0	28.2	15.8
<i>Model 2</i>	-23323.4	28.1	22.4
<i>Model 3</i>	-21474.6	27.8	18.8
<i>Model 4</i>	-21425.8	28.3	16.0
<i>Model 5</i>	-23402.9	30.0	16.1
<i>Model 6</i>	-23209.7	28.9	16.2
<i>Model 7</i>	-24625.4	28.3	16.1

NOTE: Model 1 (CM-Poisson model)

Model 2 (Poisson model)

Model 3 (Geometric model without covariates)

Model 4 (CM-Poisson model without covariates)

Model 5 (CM-Poisson model without heterogeneity and covariates)

Model 6 (CM-Poisson without without day of week effects)

Model 7 (Log-normal model)

Scenario 1 (Predictions on the last observation of the estimated sample)

Scenario 2 (Predictions on a holdout sample of 500 households)

^aLog Marginal Density

^bMean Absolute Deviation (in days)

FIGURE CAPTIONS

Figure 1

Frequency of inter-delivery times for the 529 households in the estimation sample. The horizontal axis is measured in days and the vertical axis is the frequency of purchases.

Figure 2

Scatter plot of the average inter-delivery time vs. the average purchase quantity for each of the 529 customers. The horizontal axis is the average inter-delivery time in days and the vertical axis is the average purchase quantity in dollars.

Figure 3

A schematic representation of the four MCMC samplers. Variables in boxes are the parameters to be estimated, variables in ellipses are the co-variates and variables without boxes/ellipses are the prior parameters.

Figure 4

The cumulative hazard functions for the CM-Poisson and the Poisson models. The horizontal axis is the time in weeks and the vertical axis is the cumulative hazard function. These functions have been computed from the survivor curve which was produced by integrating over the heterogeneity distribution over λ_{ij} .

Figure 1

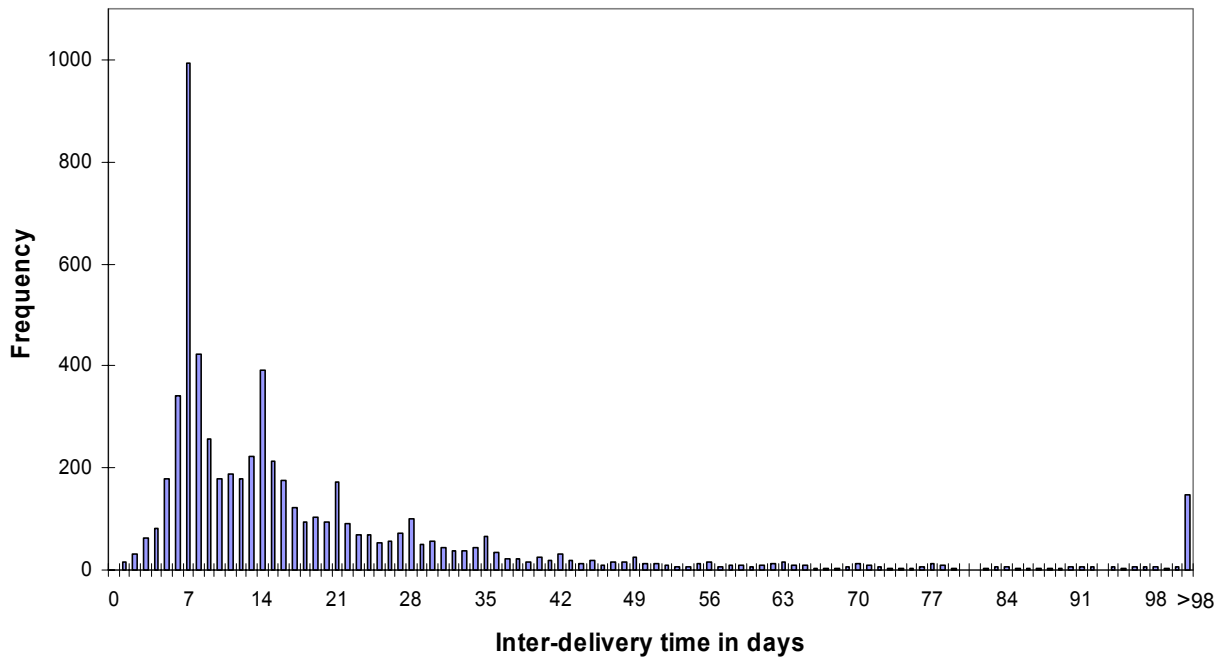


Figure 2

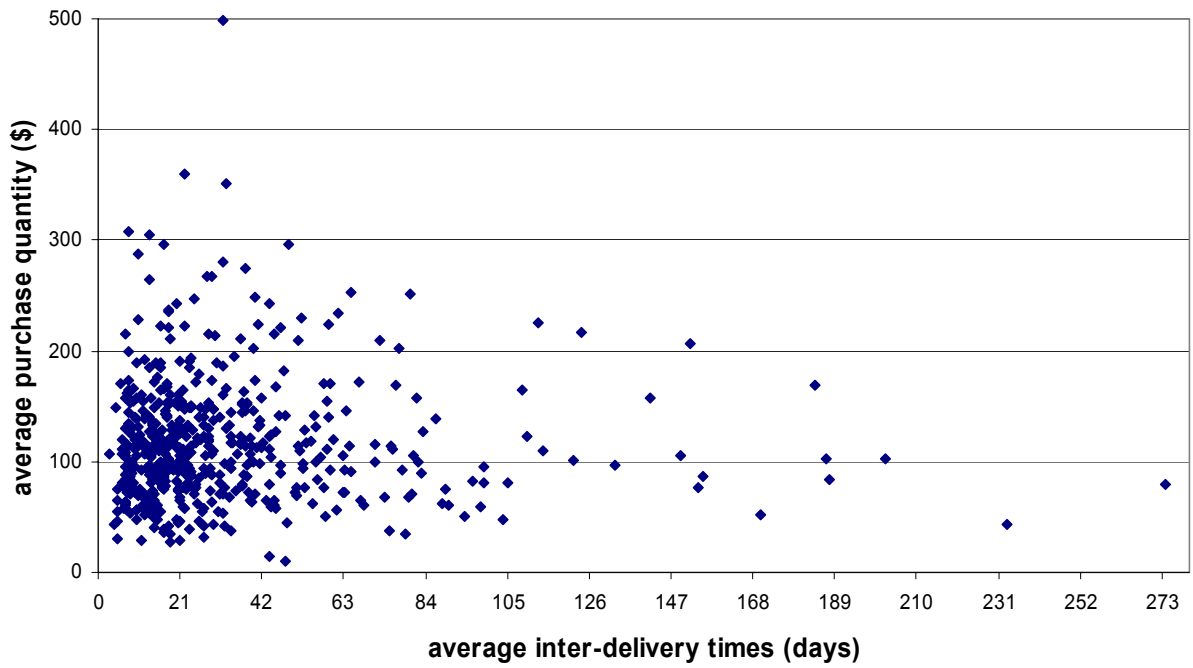
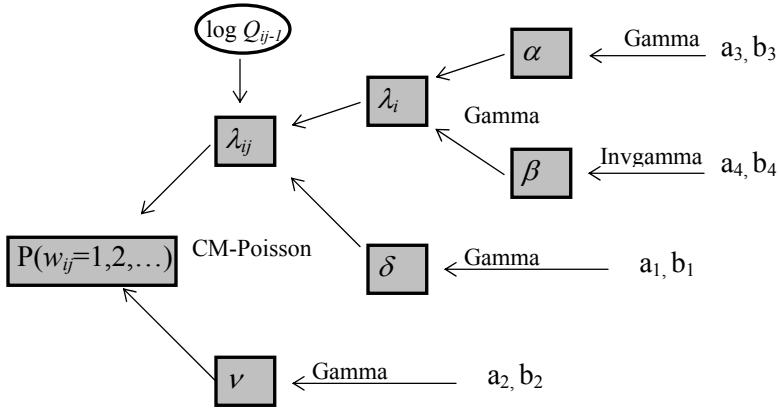


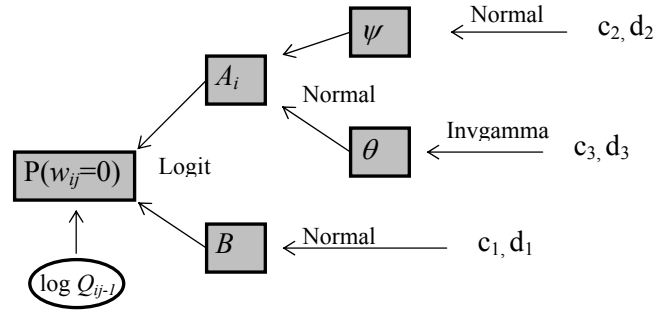
Figure 3

Sampler 1: Inter-delivery Weeks 1 and Greater

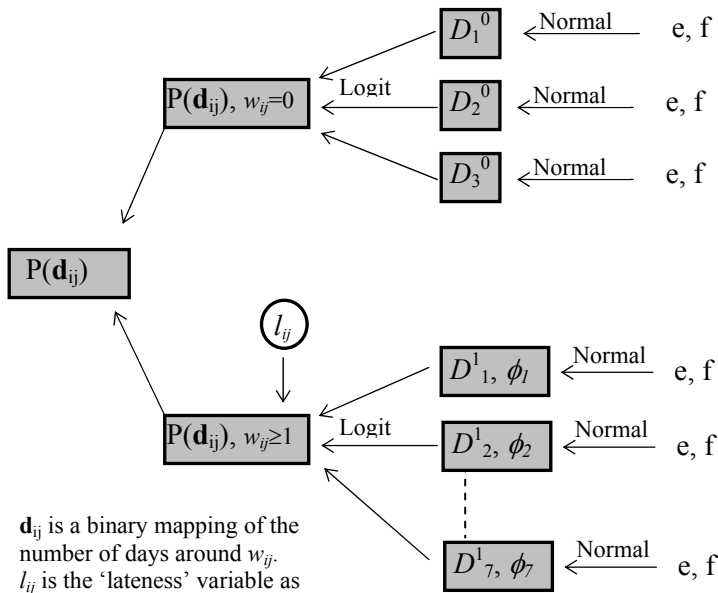


w_{ij} is the inter delivery time in weeks rounded off to the nearest week.
 Q_{ij-1} is the purchase quantity (in dollars) on the $(j-1)^{\text{th}}$ occasion.

Sampler 2: Inter-delivery Week 0

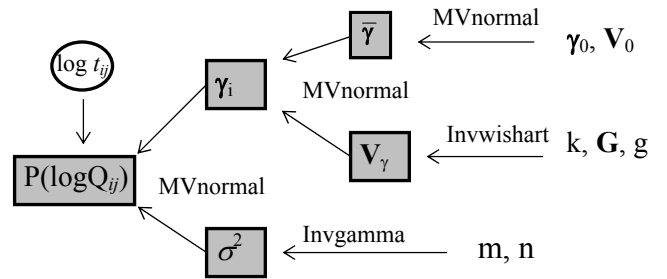


Sampler 3: Inter-delivery Days



\mathbf{d}_{ij} is a binary mapping of the number of days around w_{ij} .
 l_{ij} is the 'lateness' variable as defined in (10)

Sampler 4: Purchase Quantity



t_{ij} is the inter delivery time in days.

Legend

- Variables in boxes are the parameters to be estimated
- Variables in ellipses are the co-variates
- Variables without boxes/ellipses are the prior parameters

Figure 4

