2004

Capturing Evolving Visit Behavior in Clickstream Data

Wendy W. Moe

Peter S. Fader
University of Pennsylvania

Recommended Citation

Moe, W. W., & Fader, P. S. (2004). Capturing Evolving Visit Behavior in Clickstream Data. *Journal of Interactive Marketing, 18* (1), 5-19. http://dx.doi.org/10.1002/dir.10074

Capturing Evolving Visit Behavior in Clickstream Data

Abstract
Many online sites, both retailers and content providers, routinely monitor visitor traffic as a useful measure of their overall success. However, simple summaries such as the total number of visits per month provide little insight about individual-level site-visit patterns, especially in a changing environment such as the Internet. This article develops an individual-level model for evolving visiting behavior based on Internet clickstream data. We capture cross-sectional variation in site-visit behavior as well as changes over time as visitors gain experience with the site. In addition, we examine the relationship between visiting frequency and purchasing propensity at an e-commerce site. We find evidence supporting the notion that people who visit a retail site more frequently have a greater propensity to buy. We also show that changes (i.e., evolution) in an individual's visit frequency over time provides further information regarding which customer segments are likely to have higher purchasing conversion rates.

Disciplines
Business | Business Analytics | E-Commerce | Management Sciences and Quantitative Methods | Marketing | Technology and Innovation

Capturing Evolving Visit Behavior in Clickstream Data

Wendy W. Moe and Peter S. Fader

The Wharton School
University of Pennsylvania
Marketing Department
1400 Steinberg Hall-Dietrich Hall
Philadelphia, PA 19104-6371

telephone: (215) 898-7603
fax: (215) 898-2534
e-mail: wendy@marketing.wharton.upenn.edu

March 2000

The authors would like to thank Media Metrix, Inc. for generously providing the data needed for this paper as well as Eric Bradlow and Bruce Hardie for their insightful and valuable comments.

Capturing Evolving Visit Behavior in Clickstream Data

Abstract:
Many online retailers monitor visitor traffic as a measure of their stores' success. However, summary measures such as the number of hits per month provide little insight into individual consumers' behavior. Additionally, behavior may evolve over time, especially in a changing environment like the Internet. Understanding the nature of this evolution provides valuable knowledge that can influence how a web store is managed and marketed.

This paper develop an individual-level model for store visiting behavior based on Internet clickstream data. We capture cross-sectional variation in store-visit behavior as well as changes over time as consumers gain experience with the store. That is, as consumers make more visits to a site, their latent rate of visit may increase, decrease, or remain unchanged as in the case of static, mature markets. So as the composition of the consumer population changes (e.g., as consumers mature or as large numbers of new and inexperienced Internet shoppers enter the market), the overall degree of consumer heterogeneity that each store faces may shift.

We also examine the relationship between visiting frequency and purchasing propensity. Previous studies suggest that consumers who shop frequently may be more likely to make a purchase on any given shopping occasion. As a result, frequent shoppers often comprise the preferred target segment. We find evidence supporting the fact that people who visit a store more frequently are more likely to buy. However, we also show that changes (i.e., evolution) in an individual's visit frequency over time provides further information regarding which consumer segments are more likely to buy. Rather than simply targeting all frequent shoppers, our results suggest that a more refined segmentation approach that incorporates how much an individual's behavior is changing could more efficiently identify a profitable target segment.

1. Introduction

Many online retailers monitor visitor traffic as a measure of their stores' success. Unlike the bricks-and-mortar environment where only purchases are easily observable, the Internet allows marketers to better understand their customers by also analyzing visiting patterns. It is particularly important to gain a better understanding of online visiting behavior since over 70% of Internet retailers experienced a less than 2% conversion rate (purchase transactions per visit) in 1999 (Forrester 1999). This stands in stark contrast to the grocery store industry that marketers have studied in depth, where virtually every visit to the store is associated with a purchase. With such low conversion rates at online stores, understanding consumers' visiting behavior is a first step toward better understanding their purchasing process and the role that visits play in this process.

Though many online retailers monitor *overall* visitor traffic at their store sites, commonly used summary measures such as the total number of hits per month provide little insight into an individual consumer's behavior. For example, if the total number of visits to a store site is increasing from month to month, is the store necessarily successful? The optimistic manager would say that the store is growing by leaps and bounds. However, upon further thought, it becomes less clear whether that is the case. It is possible that the store is attracting a large number of first-time users while existing customers may be visiting less frequently over time or even dropping out completely. If such a pattern were to continue, future prospects for the store would appear less promising, especially when the arrival of new users inevitably begins to taper off. All aggregate measures (hits, page views, average time spent at site) cannot disentangle such behavioral patterns. One objective of this paper, therefore, is to offer an individual-level model of store visiting frequency, based on Internet clickstream data, that will provide more useful information about consumer behavior than such commonly used summary measures.

Evolving Visit Behavior.

A second objective of this paper is to explicitly account for evolving behavior in an individual-level model of consumer visiting. Most quantitative models of consumer purchasing assume that behavior is unchanging over time (see Morrison and Schmittlein 1988). However, since these models are being tested in stable and mature markets, such an assumption may indeed hold. But in many new markets that are still evolving, consumer behavior is known to be nonstationary (Fader and Hardie 1999). In other words, an individual's shopping behavior changes as she adapts to the new environment. The model presented in this paper will relax the assumption of stationarity. More specifically, the evolutionary component of the model allows consumers to return to the store either more or less frequently as they gain experience, while also allowing behavior to remain unchanged in some instances to accommodate the static behavior often seen in mature markets.

In the relatively new and fast-paced Internet environment, it is particularly important to address the issue of evolving behavior as consumers are continually updating their behavior, and web retailers must adapt to keep up with their customers. For example, studies have shown that as consumer knowledge and familiarity increase over time, the extent of search a consumer undertakes changes, either increasing or decreasing depending on the situation (see Alba and Hutchinson 1987, Johnson and Russo 1984, Park, Iyer, and Smith 1989). Typically, increased store knowledge and familiarity lead to more efficient search behavior. This increased shopping efficiency may have one of two effects on future store visiting behavior. One, the amount of explicit search required to make a purchase decision decreases as consumers have more internal knowledge from which to draw (Bettman 1979, Johnson and Russo 1984, Park, Iyer, and Smith 1989). This may lead to *less* frequent store visits as the consumer adapts to the shopping situation. On the other hand, Johnson and Russo (1984) have also shown that more knowledgeable consumers will search more since they can search more efficiently. As a result, store visits may become *more* frequent over time for an individual shopper.

These theoretical results suggest that as a consumer repeatedly visits a store and becomes more familiar and knowledgeable with the process, future visits may become either more or less frequent. Though other researchers have hypothesized and tested the direction of this behavioral change under various circumstances, it is not the objective of this paper to do so. Rather, our objective is to develop a flexible model that will accommodate varying magnitudes and directions of the behavioral change and offer a method to characterize the nature of this evolution.

From our evolving visit model, we can estimate how likely (and when) a given consumer will return to the store as she gains experience with a website. Do intervisit times tend to speed up or slow down over a consumer's history, and how do these changes vary across people? Answers to these questions will give us the ability to forecast future store visits in order to better anticipate and manage website traffic. We will show that our evolving model of visiting behavior forecasts significantly better than an equivalent static model. Additionally, the evolutionary component of the model will offer useful diagnostics that will help shed light on other aspects of online shopping behavior.

As we better describe the consumers in terms of their visiting behavior, we will also discuss the relationship that visiting frequency has with purchasing propensity. Previous studies suggest that consumers who shop frequently may be more likely to make a purchase on any given shopping occasion (Janiszewski 1998, Jarboe and McDaniel 1987, Roy 1994). As a result, frequent shoppers are often the preferred target segment. We find evidence consistent with this notion that people who visit a store more frequently are more likely to buy. However, we also show that changes (i.e., evolution) in an individual's visit frequency over time provides even better information regarding which customers (and customer segments) are more likely to buy. Rather than simply targeting all frequent shoppers, our results suggest that a more refined segmentation approach that incorporates how much an individual's behavior is changing can more efficiently identify profitable customers for targeting purposes.

In the next two sections, we will develop the model and address some of the key estimation issues that arise from the model. We then describe the clickstream data that we will be using. In §5, we will present the results of the model when applied to two leading online retailers and briefly discuss some of the managerial implications of the results. We will also validate the model by demonstrating its forecasting ability. Finally, in §6, we will illustrate how purchasing behavior varies across consumers as a function of their latent visit rate as well as changes in this rate over time.

2. Model Development

To understand an overall pattern of store visits, let us imagine that each consumer tends to return to a store at a latent rate inherent to that individual. *When* that individual will visit the store next is driven largely by this rate of visit. Additionally, since consumers are heterogeneous, this rate of visit varies from person to person. Some consumers may visit the store fairly frequently while others may not. But in addition to varying rates of visit across individuals, behavior may also change over time for a given individual. As consumers mature, perhaps as a result of increased knowledge and experience, their behavior may evolve thereby changing their rates of visit over time.

To capture the processes described above, our model has three main components:

(1) A timing process governing an individual's rate of visiting,

(2) A heterogeneity distribution that accommodates differences across consumers, and

(3) An evolutionary process that allows a given individual's underlying visit rate to change from one visit to the next.

Timing process with heterogeneity

As a very robust starting point, consumer visiting can be modeled as an exponential-gamma (EG) timing process. That is, each individual's intervisit time is assumed to be exponentially distributed governed by a rate, λ_i.[1] Furthermore, these individual rates of visit vary across the

[1]Alternative timing distributions, such as the Erlang, were also examined but consistently performed worse than the exponential.

population. This heterogeneity can be captured by a gamma distribution with shape parameter, *r*, and scale parameter, α. These distributions are given by the following two densities:

$$f(t_{ij};\lambda_i) \;=\; \lambda_i e^{-\lambda_i(t_{ij}-t_{h(j-1)})} \qquad \textit{and} \qquad g(\lambda_i;r,\alpha) \;=\; \frac{\lambda_i^{r-1}\alpha^r e^{-\alpha\lambda_i}}{\Gamma(r)} \tag{1}$$

where λ_i is individual *i*'s latent rate of visit, t_{ij} is the day when the j^{th} repeat visit occurred, and t_{i0} is the day of their initial visit. For a single visit occasion, this leads to the following familiar exponential-gamma mixture model:

$$f(t_{ij};r,\alpha) \;=\; \int_0^\infty f(t_{ij};\lambda_i)\cdot g(\lambda_i;r,\alpha)\,d\lambda \;=\; \frac{r}{\alpha}\left(\frac{\alpha}{\alpha+(t_{ij}-t_{i(j-1)})}\right)^{r+1} \tag{2}$$

While the exponential-gamma may be an excellent benchmark model, it fails to capture nonstationarity over time. To account for nonstationarity, extensions of this model are described next.

Evolving Behavior

A principal objective of this paper is to examine the issue of nonstationary or evolving visiting behavior. Sabavala and Morrison (1981) incorporated nonstationarity by introducing a renewal process into a probability mixture model in accordance with the "dynamic inference" framework first set out by Howard (1965). Sabavala and Morrison applied this model to explain patterns of advertising media exposure over time; further applications of a similar type of renewal-process approach can be seen in Fader and Lattin 1993 as well as Fader and Hardie 1999. These papers capture nonstationarity by assuming a probabilistic renewal process in which customers occasionally discard their old rate parameters and draw new ones from the original heterogeneity distribution. While this is a very powerful, effective way to capture longitudinal changes at the individual level, it is not consistent with the type of gradual, evolutionary behavioral changes that are likely to occur from visit to visit. That is, we expect individual behaviors to update in a smooth way as opposed to the larger, more abrubt changes that would correspond to an entirely new rate parameter (albeit from the same distribution).

In contrast, our behavioral assumption is that consumers' underlying rates of visiting are continually and incrementally changing from one visit to the next. As individuals adapt to and gain experience with the new retail environment, they may return to the store at a more frequent rate, a less frequent rate, or perhaps at the same rate for the next visit. By assuming that each individual will update her latent rate, λ_i, after each visit, a very simple way to specify this updating process is as follows:

$$\lambda_{i(j+1)} = \lambda_{ij} \cdot c \qquad \textbf{(3)}$$

where λ_{ij} is the rate associated with individual i's j^{th} repeat visit and c is a multiplier that will update this rate from one visit to the next. If the updating multiplier, c, equals one, consumer visiting is considered unchanging, and the stationary exponential-gamma would remain in effect. But if c is greater than one, consumers are visiting more frequently as they gain experience, and if c is less than one, consumers are visiting less frequently as they gain experience.

However, using a constant multiplier to update the individual λ's would be a very restrictive (and highly unrealistic) way of modeling evolutionary behavior in a heterogeneous environment. A more general approach is to replace the scalar multiplier, *c*, with a random variable c_{ij} in order to acknowledge that these updates can vary over time and across consumers. Each individual visit will lead to an update that may increase, decrease, or retain the previous rate of visit, depending on the stochastic nature of the updating multiplier.

To generalize (3) in this manner, we assume that these probabilistic multipliers, c_{ij}, arise from a gamma distribution, common across individuals and visits, with shape parameter *s* and scale parameter β. This gamma distribution essentially describes the nature of the behavioral evolution faced by a given store. The updated $\lambda_{i(j+1)}$ then becomes a product of two independent gamma-distributed random variables: the previous rate, λ_{ij}, and the multiplier, c_{ij}. The overall model, therefore, uses four parameters to simultaneously capture cross-sectional heterogeneity and evolving visiting behavior: two parameters (*r* and α) govern the gamma distribution that describes the initial heterogeneity in visiting rates, and another two parameters (*s* and β) govern the gamma distribution that describes the updating process. This is the entire model specification.

Regardless of whether the multiplier is increasing ($c_{ij}>1$) or decreasing ($c_{ij}<1$) a particular visit rate at a particular point in time, we expect that an individual's value of λ will evolve relatively slowly over time. This suggests that the updating gamma distribution, $u(c_{ij}; s, \beta)$, should have a mean fairly close to 1.0 but should also allow for more extreme increases or decreases in λ at any given update opportunity. The spread of this updating distribution is directly tied to the magnitude of the *s* and β parameters. As both of these parameters become larger, the distribution begins to degenerate towards a spike located at s/β. Taken to the extreme (i.e., *s* and β get extremely large), this model would then collapse into the deterministic updating model (3) with c= s/β.

Finally, another interesting characteristic of the updating distribution is that it allows for customer attrition, since the gamma distribution can yield a draw of c_{ij} extremely close to 0. When this situation arises, the consumer effectively drops out and is unlikely to return to the site. Such attrition may be very common for websites and has been the centerpiece of other types of models in this general methodological area (Fader and Hardie 1999; Schmittlein, Morrison, and Colombo 1989). The fact that we can accommodate attrition in such a simple, natural manner is an appealing aspect of the proposed modeling approach.

3. Likelihood Specification

When estimating the ordinary (stationary) exponential-gamma model, there are two ways of obtaining the likelihood function for a given individual. The usual approach is to specify the individual-level likelihood function, conditional on that person's (unobserved) value of λ_i. This likelihood is the product of J_i exponential timing terms, where J_i is the number of repeat visits made by household i, plus an additional term to account for the right-censoring that occurs between that customer's last arrival and the end of the observed calibration period (at time T):

$$L_i|\lambda_i = \lambda_i e^{-\lambda_i(t_{i1}-t_{i0})} \cdot \lambda_i e^{-\lambda_i(t_{i2}-t_{i1})} \cdot \quad \cdot \lambda_i e^{-\lambda_i(t_{iJ_i}-t_{i(J_1-1)})} \cdot e^{-\lambda_i(T-t_{iJ_i})} \tag{4}$$

To get the unconditional likelihood we then integrate across all possible values of λ, using the gamma distribution as a weighting function:

$$L_i|r,\alpha = \int_0^\infty L_i|\lambda_i \cdot gamma(\lambda_i; r, \alpha)\, d\lambda_i \qquad \textbf{(5)}$$

where gamma(λ_i; r, α) denotes the gamma distribution as shown in (1).

This yields the usual exponential-gamma likelihood, which can be multiplied across the N households to get the overall likelihood for parameter estimation purposes:

$$L = \prod_{i=1}^{N} \frac{\Gamma(r+J_i)}{\Gamma(r)} \left(\frac{\alpha}{\alpha + T - t_{i0}}\right)^r \left(\frac{1}{\alpha + T - t_{i0}}\right)^{J_i} \qquad \textbf{(6)}$$

An alternative path that leads to the same result is to perform the gamma integration separately for each of the J_i+1 exponential terms, and then multiply them together at the end. This involves the use of Bayes Theorem to refine our "guess" about each individual's value of λ_i as each arrival occurs. Specifically, it is easy to show that if someone's first visit occurs at time t_{ij}, then:

$$g(\lambda_{i2}|arrival\ at\ t_{i1}) = gamma(r+1, \alpha + t_{i1} - t_{i0}) \qquad \textbf{(7)}$$

The gamma distribution governing the rate of visit for subsequent arrivals follows:

$$g(\lambda_{i(j+1)}|arrival\ at\ t_{ij}) = gamma(r+j, \alpha + t_{ij} - t_{i0}) \qquad \textbf{(8)}$$

Using this logic, we can re-express the likelihood as a series of separate EG terms:

$$L = \prod_{i=1}^{N} \prod_{j=1}^{J_i} \left(\frac{r+j+1}{\alpha + t_{i(j-1)} - t_{i0}} \right) \left(\frac{\alpha + t_{i(j-1)} - t_{i0}}{\alpha + t_{ij} - t_{i0}} \right)^{r+j} \cdot S(T - t_{iJ_i})$$

$$\text{where} \quad S(T - t_{iJ_i}) = \left(\frac{\alpha + t_{iJ_i} - t_{i0}}{\alpha + T - t_{i0}} \right)^{r+J_i} \qquad \textbf{(9)}$$

Which collapses into the same expression as (6).

When we introduce the nonstationary updating distribution, the multipliers (c_{ij}) change the value of λ_i from visit to visit, thereby requiring us to use the sequential approach given in (9) to derive the complete likelihood function. We need to capture two forms of updating after each visit: one due to the usual Bayesian refinement process (which is associated with stationary behavior given by (8)) and the other due to the effects of the stochastic evolution process. Therefore, the distribution of visiting rates at each repeat visit level is the product of two gamma distributed random variables – one associated with the updating multiplier and one capturing the previous visiting rate. For the case of a observing household *i* making her *j*th repeat visit at time t_{ij}:

$$G(\lambda_{i(j+1)} | \text{arrival at } t_{ij}) = gamma(r+j, \alpha + t_{ij} - t_{i0}) \cdot gamma(s, \beta) \qquad \textbf{(10)}$$

One issue with this approach is that the product of two gamma random variables does not lend itself to a tractable analytic solution. However, there is an established approach (see, e.g., Kendall and Stuart 1977, p. 248) suggesting that the product of two gamma distributed random variables can itself be approximated by yet another gamma distribution, obtained by multiplying the first two moments about the origins of the original distributions:

$$m_1^{(\lambda_{i(j+1)})} = m_1^{(\lambda_{ij})} \times m_1^{(c_{ij})}$$

$$\text{and} \qquad \textbf{(11)}$$

$$m_2^{(\lambda_{i(j+1)})} = m_2^{(\lambda_{ij})} \times m_2^{(c_{ij})}$$

As shown in Appendix A, this moment-matching approximation, used in conjunction with Bayesian updating, allows us to recover the updated gamma parameters that determine the rate of visit, λ_{ij}, for household i's j^{th} repeat visit as follows:

$$r(i,j+1) = \frac{\left[r(i,j)+1\right]\cdot s}{\left[r(i,j)+2\right]\cdot(s+1) - \left[r(i,j)+1\right]\cdot s} \qquad \textbf{(12)}$$

$$\alpha(i,j+1) = \frac{\left[\alpha(i,j)+t_{ij}-t_{i(j-1)}\right]\cdot\beta}{\left[r(i,j)+2\right]\cdot(s+1) - \left[r(i,j)+1\right]\cdot s} \qquad \textbf{(13)}$$

where $r(i, 1)$ and $\alpha(i, 1)$ are equal to the initial values of r and α as estimated by maximizing the likelihood function specified in (8).

We performed 20 simulations to verify the accuracy of using such a moment-matching approximation. In each simulation, we first generated 1000 random draws from a gamma distribution with randomly determined shape and scale parameters to represent initial λ values. Then, a matrix of updating multipliers were also simulated for a series of five updates or five future repeat visits. Each 1000 x 5 matrix was generated by taking draws from a gamma distribution, again with randomly determined shape and scale parameters, where columns one through five represented the updates after one to five visits. The updated λ series after five repeat visits was calculated using two methods (1) direct (numerical) multiplication of the 1000 initial λ's and the five updating series or (2) randomly drawing 1000 values from the distribution resulting from the moment-matching approximation across all five updates. A Kolmogorov-Smirnov test of fit indicated that, for each of the 20 simulations, the distribution of values resulting from the moment-matching approximation is not significantly different from that resulting from the direct multiplication of these random variables. Therefore, we are confident that the moment-matching approximation accurately captures the gamma distributed updating process we wish to model.

After incorporating the evolution process into our model, the likelihood function to be maximized follows:

$$L = \prod_{i=1}^{N} \prod_{j=1}^{J_i} \left(\frac{r(i,j)}{\alpha(i,j)} \right) \left(\frac{\alpha(i,j)}{\alpha(i,j) + t_{ij} - t_{i(j-1)}} \right)^{r(i,j)+1} \cdot S(T - t_{iJ_i}) \tag{14}$$

where $r(i, j)$ and $\alpha(i, j)$ are defined in equations (12) and (13) while the survival function, $S(T\text{-}t_{ij})$, is defined as:

$$S(T - t_{ij}) = \left(\frac{\alpha(i, J_i + 1)}{\alpha(i, J_i + 1) + T - t_{iJ_i}} \right)^{r(i, J_i + 1)} \tag{15}$$

For the special case in which behavior is not evolving and the nonstationary updating distribution degenerates to a spike at 1.0 (i.e., $s = \beta = M$, where $M \to \infty$), then this equation collapses down exactly to the ordinary (stationary) exponential-gamma model.

4. Data

We apply the models described in the previous section to clickstream data collected by Media Metrix, Inc. Media Metrix maintains a panel of approximately 10,000 households whose Internet behavior (and in fact, all computer behavior) is recorded, pageview by pageview, over time. Participating households install Media Metrix software on their personal computers. While panelists surf the Internet, the software runs in the background and records the date, time, and duration of each and every page being viewed. The computer automatically uploads this detailed data to Media Metrix on a periodic basis.

For our purposes, we are interested in the dates of the visits each household makes to a given store site. Any session in which the web user views a URL with a particular online store's domain name is considered a visit to that store. To consolidate the data just a bit, we aggregated visits to the daily level. For example, if a given household were to visit a particular store multiple times in a single calendar day, we would encode that behavior as just one visit for the day when the session began. Since we are interested in the timing and frequency of repeat visits to a store, our dataset describes each household as a sequence of days when visits were made. All households that have visited the store of interest at least once during the observation period were included in this dataset.

To illustrate our model, we use data from March 1, 1998 to October 31, 1998 for two online store sites. One is a leading online bookstore while the other is a popular online CD store. The bookstore attracted 4,379 unique visitors to its site during this eight-month period totaling 11,263 visits, while the CD store had 1,670 visitors making 3,616 visits. Figure 1 provides a histogram of the number of visits each household made to each store.

[***Figure 1.*** *Histogram of Household Visits*]

More interesting than this static snapshot of the number of visits across the entire time period, however, is a look at how the number of visits per period changes over time. Table 1 shows the number of unique visitors and the total number of visits for the first four months in the dataset versus the second four months for both the bookstore and the CD store. Not only do these measures show an increase in the total number of visits over time, but there also appears to be an increase in the number of visits per visitor.

[***Table 1.*** *Summary of Visit Data over Time*]

These aggregate summary statistics suggest that consumers are visiting more frequently over time. On the surface, this seems like great news for the store managers. However, as we suggested at the outset of the paper, these numbers may be misleading – many individuals may be experiencing a slowdown in their visit rates, but an influx of new visitors (with relatively high visit rates) in the latter period might be masking these dynamics. This example provides a strong motivation for the type of individual-level dynamic model that we have proposed here. In the next section we examine the empirical evidence that will shed light on the actual behavioral patterns at play here.

5. Model Results

Before estimating the evolving visit model developed in §3, we first examine the static exponential-gamma timing model as a benchmark. When the static, two-parameter model is applied to the eight months of bookstore data, we find that the mean rate of visit ($E[\lambda]=r/\alpha$) is 0.0112. In other words, the average intervisit time ($1/\lambda$) is 89.3 days, which is fully consistent with the summary statistics mentioned earlier. But beyond their ability to capture the mean of the heterogeneous visiting process, the model parameters also provide useful information about the nature of the distribution of visit rates across the population. With a shape parameter of 0.483 and a scale parameter of 42.955, the distribution of consumer visiting rates can be described by the gamma distribution in Figure 2. This distribution has a large proportion of the consumer population with very low rates of visit. The median rate, according to this model, is 0.005, corresponding to an intervisit time of 200 days.

[***Figure 2.*** *Gamma Distribution of Visiting Rates for Stationary EG Model*]

A principal reason for these high intervisit times is the fact that the stationary model does not allow consumers to drop out and never return. As a result, a consumer that has actually dropped out would be seen by the model as having a very slow visiting rate, since she would not have yet returned to the store by the end of the observation period. The evolving visit model, however, allows for dropout (as well as evolving rates among visitors) and therefore provides more reasonable estimates of intervisit times.

In Table 2 we contrast the parameter estimates and fit statistics for the static EG model with those from our four-parameter model of evolving visiting behavior. Not only does the latter model fit the data better, but it also has more intuitively appealing results. While the basic shape of the gamma distribution for initial visit rates (shown in Figure 3a) may appear to be similar to that of the static EG model, it is less dominated by low-frequency shoppers, leading to a substantially lower mean intervisit time (52 days, E[λ] = 0.019). Likewise, the median intervisit time shrinks to 167 days (median λ =0.006). These differences reflect the fact that dropout – or other types of evolution – can take place as the consumer becomes more familiar with the site.

[***Table 2.*** *Model Results for Bookstore*]

According to the evolutionary model, the mean update for any given visit (s/β) is very close to one (0.998) suggesting, perhaps, that it is a fairly stationary process. However, a closer look at the distribution (see Figure 3b) shows that there is significant variance about this mean. Though the mean update is close to one, the distribution is quite skewed. With a median value of c_{ij}=0.858, consumers decrease their shopping frequency over 85% of the time. That is, from visit to visit, consumers tend to return to the store at slower rates. The implications of these results are in stark contrast to the measures summarized in Table 1 that implied increased visiting frequency over time.

[***Figure 3.*** *Evolving Visit Model Distributions for Bookstore Data*]

Though other models have acknowledged the issue of nonstationarity, they have focused primarily on dropout (Eskin 1973, Kalwani and Silk 1980, Schmittlein, Morrison, and Columbo 1987). These models allow for individuals to make several purchases, become disenchanted, and never purchase again. To test if the evolving visit model is capturing evolving behavior over time in addition to a dropout phenomenon, we also estimated an exponential-gamma model with a dropout component similar to that specified by Eskin 1973 and Fader and Hardie (1999).

In the EG model with dropout, the probability of visiting given that you are an active visitor is modeled as an exponential-gamma process. However, the probability of being an active visitor after the j^{th} visit, π_j, is determined by the following:

$$\pi_j = \phi(1 - e^{-\theta j}) \tag{16}$$

where ϕ is the long run probability of a consumer dropping out, and θ is the rate at which the dropout rate approaches this long run probability. Though the EG model with dropout provides a significant improvement in fit over the stationary EG model (LL = -33,804.7), it does not approach the performance of the evolving visit model which has the same number of parameters. This suggests that the evolving visit model is capturing a phenomenon in addition to just dropout.[2]

Validation

While we have discussed the fact that the evolving visit model fares well on a relative basis compared with various benchmark models, we have yet to show that it performs sufficiently well on an absolute basis. In this section, we will validate the evolving visit model by examining the accuracy of longitudinal forecasts. Because the evolving visit model relies on an approximation

[2]We also tested several nested models that allowed for a constant update after every visit (i.e., equation 3), both with and without the dropout process. None of these models came close to the proposed evolving visit model in terms of fit or forecasting performance.

(11) to specify and estimate the model, we need to perform simulations to generate data for tracking/forecasting purposes. This is a straightforward and computationally efficient task. For each iteration of the simulation, we create a simulated panel that matches the actual panel in terms of its size and the distribution of its initial visit times. We then generate a sequence of repeat visits using the parameter estimates from the model. This requires us to maintain a time-varying vector of λ's for each household, which starts with random draws from the initial (r, α) gamma distribution, and then gets updated using the (s, β) gamma distribution after each simulated exponential arrival occurs. We continue this process until every simulated household gets past the tracking/forecasting horizon of interest to us. It is then a simple matter to count up the number of visits on a week-by-week basis for each iteration of the simulation. We then average across 1000 iterations to generate the tracking and forecasting plots. Using the MATLAB programming language, each of these iterations takes only a few seconds on a standard PC, and we see very consistent convergence properties after a few dozen iterations.

Before creating the forecasts, we re-estimate both models (stationary and evolving EG) using only the first half (i.e., four months) of the dataset. (It is worth noting that the evolving model parameters are quite robust to this changing calibration period, while the stationary model has a noticeably higher visit rate over the shorter period – clear evidence of the slowdown discussed earlier). To generate the forecasts for the evolving visit model, we use the simulation procedure described above. For the stationary EG model, the expected number of repeat visits per week can be calculated directly as follows:

$$E[repeat\ \ visits_w] \; = \; N_w\, t \left(\frac{r}{\alpha} \right) \qquad \textbf{(17)}$$

where N_w is the number of eligible repeat visitors in week w and t is the time period of interest, i.e., seven days in this case. Figure 4 shows cumulative forecasts as well as actual visits for the bookstore site.

[***Figure 4.*** *Forecasts of Repeat Visits*]

Both models seem to track the data quite well over the initial four-month calibration period. However, as we enter the forecasting period, the stationary EG model begins to diverge, ultimately overpredicting by 37% for the bookstore at the end of the eight month period. It overestimates the number of visits per week as it does not recognize that consumers are returning less frequently over time. The evolving visit model, however, forecasts quite accurately, well within 5% of the actual sales line throughout the forecast period. This is an impressive achievement and serves as a strong testimonial to the validity of the assumptions, structure, and parameter estimates associated with the proposed model.

Results for CD Store

The same set of models and analyses were also applied to the CD store data (results in Table 3). We see a remarkably similar set of patterns as in the case of the bookstore. In moving from the static EG model to the evolving specification, we see significantly shorter intervisit times, since the latter model can accommodate consumer dropout. We also see, once again, that the mean update is close to 1.0 (0.991), but with a median of 0.837, consumer shopping frequency is more likely to decrease than increase after each visit. We emphasize once more that these results contradict the summary statistics from Table 1, which seemed to imply that shopping frequency is increasing from one visit cycle to the next.

[***Table 3.*** *Model Results for CD Store*]

Other benchmark models (involving dropout and/or constant updates) proved once again to be vastly inferior to the evolving visit model. Finally, our forecast validation led to encouraging results with projected visits only 2% above the actual number at the end of the eight month period, compared to a 40% overforecast for the stationary model. While we are very encouraged by these strong initial results, we are also surprised at the degree of similarity seen for these two sites. We certainly do not want to suggest that the specific patterns captured here will generalize to other online retailers, but there should be ample motivation for future studies to find and describe a broader range of online visiting behavior.

6. Visit Frequency and Evolution: Associations with Purchasing Behavior

Studies of mall shopping behavior have shown that more frequent shoppers tend to be "recreational" shoppers – they are more involved and more motivated in the process and thus are more likely to impulse sbuy (Janiszewski 1998, Jarboe and McDaniel 1987, Roy 1994). From these studies and others, there is a wealth of evidence (theoretical and empirical) implying that more frequent shoppers are also more likely buyers at any given visit occasion. In this section, we explore this relationship between consumers' visiting patterns and their purchasing propensities. We then extend the framework to incorporate (and separate out) the effects of evolving behavior on purchasing.

As an initial test of the traditional frequency-propensity hypothesis, we first calculate each bookstore consumer's expected rate of visit, λ_i, given the evolving visit model's estimated parameters and the consumers' observed behavior during the eight-month observation period. Using equations (12) and (13), we calculate each repeat visitor's mean rate of visit at the end of the observation period as $r(i, J_i)/\alpha(h, J_i)$. Across the 2098 repeat visitors to the bookstore, the median expected visit rate at the end of our time period was 0.0349 or an intervisit time of 29 days.

Additionally, we calculate each consumer's purchasing propensity by dividing the number of visits during which a purchase occurred by the total number of visits made by that individual. The average conversion rate across the repeat visitors was 0.139; that is, almost 14% of the visits made by these consumers were accompanied by a purchase. However, conversion rates differ for frequent shoppers, whom we define as consumers with visiting rates greater than or equal to the median (N=1062), versus infrequent shoppers, whom we define as consumers with visiting rates less than the median (N=1036). Frequent visitors have significantly higher conversion rates, averaging 16.6% compared to an average across the infrequent visitors of 11.1% (t=6.04, p<0.001)[3]. These results confirm the hypothesis that frequent visitors tend to be more valuable

[3]To account for the non-normality of these proportions, we utilize a standard arc-sine transformation of the conversion rates for all of the statistical tests discussed in this section.

customers since they are relatively more likely buyers, both on a percentage and an absolute basis.

However, the main objective of this paper is to capture – and capitalize upon – nonstationarity in consumer visiting behavior. Though consumer visit rates provide some information about the attractiveness of the visitor as a buyer, these rates change over time, and the nature of this change may have implications for the consumer's buying propensity. For example, new visitors may initially shop infrequently as they are unaccustomed to the environment. However, as they repeat visit, they begin to update their behavior. This evolving process may also be associated with greater purchasing propensity as it tends to lead to more involvement in the shopping process.

Therefore, in addition to segmenting consumers into frequent and infrequent visitors, we also characterize and segment consumers based on the extent of the behavioral evolution they have undergone during the observation period. For example, a frequent shopper who has always been a frequent shopper may be quite different from a frequent shopper who had recently evolved from being an infrequent shopper in the past.

To determine the extent of updating a consumer has undergone, we need to calculate a baseline rate of visit that would best capture their behavior if no evolution had taken place. Therefore, we calculate each individual's latent rate of visit given their observed behavior and the model results absent of any updating distribution (i.e., the value of λ associated with a stationary EG model).

The extent of updating for each consumer is the difference between their rate of visit as given by the nonstationary model and this baseline rate.[4]

The median update for repeat bookstore visitors is 0.000. A median split along this dimension divides shoppers into those who became more frequent visitors over time versus those who became less frequent visitors. We also see a difference in conversion rates (CR) along this dimension: those who increased their rate of visit were more likely to buy (N=1056, CR=15.1%) than those who decreased their rate of visit (N=1042, CR=12.7%). Once again, this difference is highly significant (t=2.68, p=0.007), suggesting that the degree of evolution is indeed related to purchase propensity.

After seeing these two strong effects, a natural question is whether each one is still present when both are taken into account simultaneously. Table 4 examines the issue by dividing repeat visitors along both dimensions into four cells, using the same median splits as before. It is interesting to note that the number of visitors in each cell is quite balanced, indicating that there is not a dominant association between frequency and updating. In other words, for every household that started with a slow visit rate and sped up towards the end of the model calibration period, there is a corresponding household that started with a very fast visit rate, but slowed down to roughly the same level by the end of the eight-month period.

---

[4]There is no significant difference in the relative position of each household in terms of its extent of evolution when the change in visiting rates is measured as an absolute difference versus a percentage change.

[***Table 4**. Bookstore Conversion Rates*]

An ANOVA on these data confirm that both main effects remain highly significant: $F_{1,2094}$ = 35.765 (p<0.001) for high vs. low frequency, and $F_{1,2094}$ = 6.473 (p=0.011) for increasing vs. decreasing frequency. Furthermore, a strong interaction ($F_{1,2094}$ = 5.035, p=0.025) emerged as well, and its presence is easily seen in Table 4. For infrequent visitors (top row), there is no meaningful difference in conversion rates, regardless of the nature of the household's updates over time. But for frequent visitors, the purchase-to-visit rate is considerably higher for those who have experienced increasing frequency. The households in the lower right cell are particularly conspicuous, with a conversion rate nearly 40% higher than the rest of the panel. This is clearly a very attractive group of repeat buyers.[5]

Table 5 presents the same analysis for the 581 households that made at least one repeat visit to the online CD store. The patterns are remarkably similar to those seen for the bookstore, with the exception of smaller sample sizes and lower conversion rates. The ANOVA model reveals significant main effects ($F_{1,577}$ = 4.044, p=0.045 for frequency, and $F_{1,577}$ = 8.810, p=0.003 for updating), with a very strong interaction ($F_{1,577}$ = 6.405, p=0.012) once again highlighting the unique nature of those households that have accelerated their visiting behavior to a relatively high rate over the course of the eight-month data collection period. The conversion rate for the households in this cell is over 60% higher than that of the three cells combined. While this

[5] In addition to this ANOVA conducted on the two dichotomous variables discussed here, we also examined equivalent regression models on the household-level data. The results are quite similar across the two datasets.

translates to only 3 percentage points on an absolute basis, this represents a very significant improvement in an industry that is just becoming aware of the critical importance of this single statistic as the most useful indicator of an online retailer's performance and future prospects (Gurley 2000).

[***Table 5.*** *CD Store Conversion Rates*]

Taken together, the analyses for these two leading online retailers suggest not only that frequent visitors are more likely buyers, but also that a more refined segmentation of visitors that incorporates *changes* in visiting behavior can identify an even more valuable segment of customers to target. This is a new and important result, worthy of management attention and further research.

7. Discussion and Conclusions

Many skeptics claim that the Internet is nothing more than a new distribution channel, and thus it should not change the way we examine customer behavior. While this may be true in certain respects, this paper highlights some of the uniquely different research perspectives that we gain from examining clickstream data. Thanks to rich new sources of data (such as Media Metrix), we can now examine behavioral phenomena that would be impossible to study using more traditional sources, such as grocery store scanner data.

The detailed, disaggregate data available to us make it possible to study the evolution of visit behavior at a retail site. The model developed here is not tailored specifically to online stores, although it might be hard to obtain the necessary data to estimate this model for a "bricks-and-mortar" retailer. For instance, many traditional retailers use some sort of tracking mechanism, e.g., a loyalty card, to capture the timing of purchases at the store, but it is hard for them to capture visits that do not involve a purchase.

We posit a behaviorally plausible – and highly parsimonious – model that allows visiting behavior to evolve gradually over time, although it also allows for more abrupt changes, such as permanent dropout from the site. And indeed, our empirical analysis reveals the fact that the average update in household visiting rates is a multiplier close to 1.0, but there is significant spread around this value. Additionally, the manner in which we implement this updating scheme – a gamma distribution to capture the different values of these multipliers – is a new methodological contribution, which merits consideration for other types of nonstationary modeling contexts.

Use of the model reveals that individual-level behavior patterns appear to contradict the perspective that one would obtain from examining the aggregate data alone. Specifically, the aggregate data seem to indicate an acceleration of visiting behavior at each of two leading e-commerce sites, yet our model parameters suggest that the typical household is experiencing a gradual slowdown in its visiting rate over time. The difference here is that an increasing number of new visitors are coming to each site over time, masking the slowdown that may be occurring for many experienced visitors. This effect could have dramatic implications for managers who neglect to examine their data at a sufficiently fine level of disaggregation.

Beyond the intuitive appeal of the model specification and its estimated parameters, we also show that it has excellent validity from an out-of-sample forecasting perspective. For both retail sites, the model tracks future visiting patterns extremely well, remaining within 5% of the actual data over the entire duration of a four-month holdout period. While this model was not constructed with forecasting in mind as a principal objective, this result certainly speaks well about its overall versatility.

Perhaps the most dramatic demonstration of the model's validity and usefulness is its ability to delineate highly significant differences in purchasing behavior across households. There is a significant amount of past literature suggesting that customers who visit a particular store frequently also tend to buy something during a relatively high proportion of those shopping trips. We provide strong confirming evidence of this hypothesis. But the evolutionary nature of our model allows us to test an equally compelling complementary hypothesis: households that experience increases in their visiting rates over time are more likely to purchase something at any given visit than those who are slowing down.

Both sites provide solid support for this new hypothesis, but also exhibit a powerful interaction that combines both of these effects. Specifically, households that combine high frequency with an upwards evolutionary trend in visiting behavior have dramatically higher conversion rates than all other households. As noted above and elsewhere (e.g., Forrester 1999) measuring and managing conversion rates is becoming increasingly crucial to e-commerce executives, so this is an important finding that merits additional investigation in later research.

Limitations and Future Research

Since this paper is among the first attempts to carefully examine online visiting behavior using clickstream data, we have deliberately kept the model as clear and simple as possible in order to highlight the chief phenomena that we have observed in these datasets. But as the types of data and methods employed here become more commonplace, we can see several extensions to the model that may be worth pursuing.

Because we have been emphasizing the importance of evolution in a new marketplace (such as online sales of books and CD's) we have paid little attention to the fact that these markets might eventually shift towards a more steady-state nature, i.e., with updates occurring less frequently and with smaller magnitudes. It is unlikely that the same distribution of updating multipliers (c_{hj}) will stay in place over a long period of time. Perhaps this distribution starts to collapse towards a spike at 1.0 as the market matures. The excellent performance of our holdout forecasts does not seem to indicate any such pattern in our datasets, but as our observation window extends to several years' worth of data in the future, we might see more benefits from such a specification.

Another way of improving on the c_{ij} distribution might be to let these multipliers vary more systematically across consumers and visits. Rather than assuming, as we do now, that each update is an independent draw from the same distribution of multipliers, we can allow the draws to be linked over time at the household level, and also allow the shape of the distribution to vary over time and across households. These extensions would require the use of computationally intensive hierarchical Bayes estimation procedures, which would then also enable the inclusion of other features, such as allowing for a correlation structure between the set of visit rate parameters and the update multipliers. But all of these extensions are well beyond the scope of this initial analysis.

Beyond these methodological issues on our “to-do” list, it is important to acknowledge the need for further process-oriented research to better explain and extend the psychological mechanisms underlying our findings concerning the relationships between conversion rates and visit dynamics. While there is ample theoretical reasoning behind the well-established frequency hypothesis, it would be useful to establish an equally solid base of explanations and controlled experimental evidence for the effect of positive vs. negative evolution, as well as the robust interaction effect we have observed.

Finally, our brief examination of conversion rates suggests that there is a need for modeling efforts that are more focused on this phenomenon by itself. While we have allowed visit behavior to evolve in our model, we have treated conversion rates as a purely static summary measure. In reality, however, the relationship between visits and purchases is likely to go through its own type of evolution. Once we have a complete understanding of the dynamic visit-purchase process, we can combine such a model with the present “visit only” model to obtain a complete picture of online buying behavior.

REFERENCES

Alba, Joseph W. and J. Wesley Hutchinson (1987), "Dimensions of Consumer Expertise," *Journal of Consumer Research*, 13 (March), 411-454.

Eskin, Gerald J. (1973), “Dynamic Forecasts of New Product Demand Using a Depth of Repeat Model,” *Journal of Marketing Research*, 10 (May), 115-129.

Fader, Peter S. and Bruce G.S. Hardie (1999), "Investigating the Properties of the Eskin/Kalwani & Silk Model of Repeat Buying for New Products," in Lutz Hildebrandt, Dirk Annacker, and Daniel Klapper (eds.), Marketing and Competition in the Information Age, Proceedings of the 28th EMAC Conference, May 11-14, Berlin: Humboldt University.

Fader, Peter S. and James M Lattin (1993), “Accounting for Heterogeneity and Nonstationarity in a Cross-Sectional Model of Consumer Purchase Behavior, *Marketing Science*, 12 (3), 304-317.

Gurley, J. William (2000), “The Most Powerful Internet Metric of All,” http://www.news.com/Perspectives/Column/0,176,403,00.html, February 21.

Howard, R.A.(1965), “Dynamic Inference,” *Operations Research*, Vol. 13 (2), 712-733.

Janiszewski, Chris (1998), “The Influence of Display Characteristics on Visual Exploratory Search Behavior,” *Journal of Consumer Research*, 25 (3), 290-301.

Jarboe, Glen R. and Carl D. McDaniel. (1987). "A Profile of Browsers in Regional Shopping Malls," *Journal of the Academy of Marketing Science*, 15 (Spring): 46-53.

Johnson, Eric and J. Edward Russo (1984), “Product Familiarity and Learning New Information,” *Journal of Consumer Research*, 11 (June), 542-550.

Kalwani, Manohar and Alvin J. Silk (1980), “Structure of Repeat Buying for New Packaged Goods,” *Journal of Marketing Research*, 17 (August), 316-322,

Kendall, Maurice G. and Alan Stuart (1977), The Advanced Theory of Statistics, 3rd edition, vol. 2, New York: Hafner.

Morrison, Donald G. and David C. Schmittlein (1988), “Generalizing the NBD Model for Customer Purchases: What Are the Implications and Is It Worth the Effort?” *Journal of Business & Economic Statistics*, 6 (2), 145-159.

Park, C. Whan, Easwar S. Iyer, and Daniel C. Smith (1989), “The Effects of Situational Factors on In-Store Grocery Shopping Behavior: The Role of Store Environment and Time Available for Shopping,” *Journal of Consumer Research*, 15 (4), 422-433.

Roy, Abhik (1994), “Correlates of Mall Visit Frequency,” *Journal of Retailing*, 70 (2), 139-161.

Sabavala, Darius J. and Donald G. Morrison (1981), “A Nonstationary Model of Binary Choice Applied to Media Exposure,” *Management Science*, 27 (6), 637-657.

Schmittlein, David C., Donald G. Morrison, and Richard Colombo (1987), “Counting Your Customers: Who Are They and What Will They Do Next?” *Management Science*, 33, 1-24.

APPENDIX A. *Moment-Matching Approximation of the Product of Two Gamma Distributions*

If *x* and *y* are two gamma distributed random variables,

x ~ Gamma (r, a)
y ~ Gamma (s, b)

then the product, $z = xy$, can be assumed to be a gamma distributed random variable

z ~ Gamma (R, A)

with shape and scale parameters, *R* and *A*, such that the first two raw moments of the *z*-distribution is the product of the moments of the *x*- and *y*-distributions.

$$m_1{}^x = \frac{r}{\alpha} \qquad m_2{}^x = \frac{r(r+1)}{\alpha^2}$$

$$m_1{}^y = \frac{s}{\beta} \qquad m_2{}^y = \frac{s(s+1)}{\beta^2}$$

$$m_2{}^z = m_1{}^x \cdot m_1{}^y = \frac{rs}{\alpha\beta} \qquad m_2{}^z = m_2{}^x \cdot m_2{}^y = \frac{r(r+1)s(s+1)}{\alpha^2\beta^2}$$

Since the first moment of the *z*-distribution, $m_1{}^z$, is R/A and the second moment, $m_2{}^z$, is $R(R+1)/A^2$, we can solve for *R* and *A* with the following two equations:

$$\frac{R}{A} = \frac{rs}{\alpha\beta} \qquad \frac{R(R+1)}{A^2} = \frac{r(r+1)s(s+1)}{\alpha^2\beta^2}$$

Therefore, the gamma distribution describing the product of two independently distributed gamma random variables has shape and scale parameters that can be calculated from the parameters of the multiplying distributions.

$$R = \frac{rs}{(r+1)(s+1)-rs} \qquad A = \frac{\alpha\beta}{(r+1)(s+1)-rs}$$

with Bayesian updating after observing one arrival at time t

$$R = \frac{(r+1)s}{(r+2)(s+1)-(r+1)s} \qquad A = \frac{(\alpha+t)\beta}{(r+2)(s+1)-(r+1)s}$$

Figure 1. Histogram of Household Visits

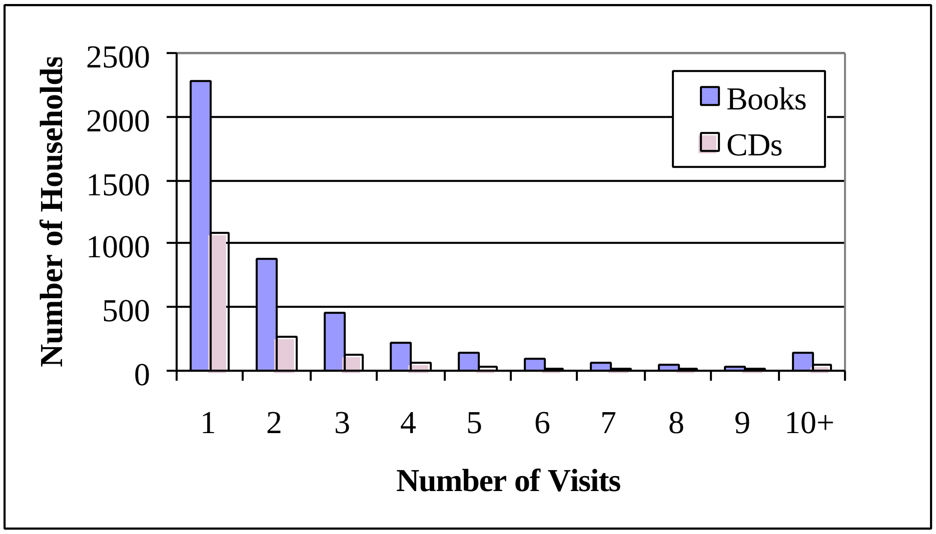


Figure 2. Gamma Distribution of Visiting Rates for Stationary EG Model

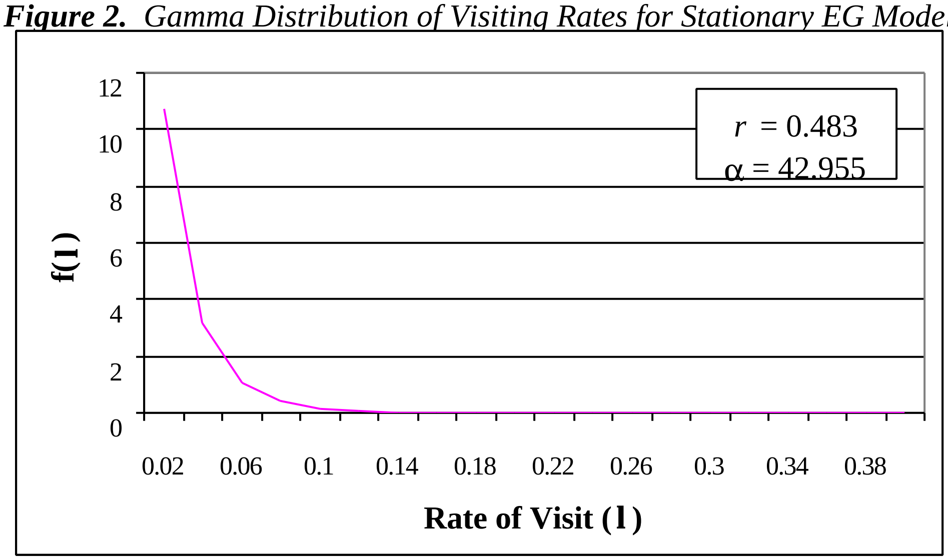
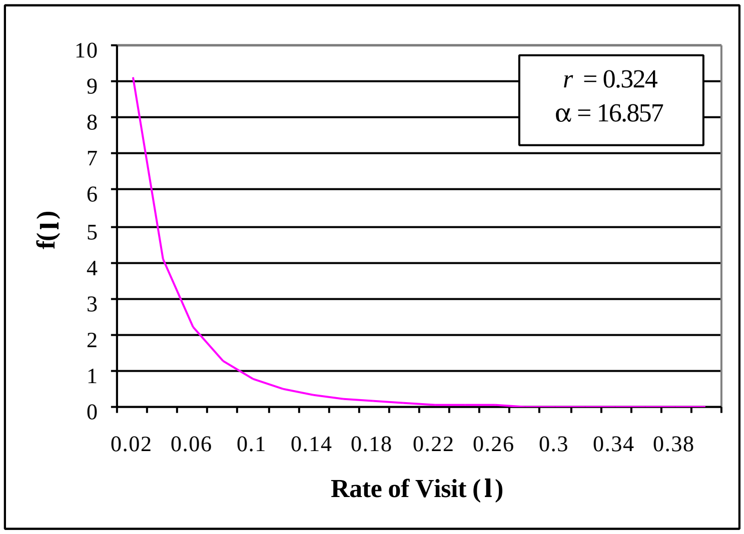


Figure 3. *Evolving Visit Model Distributions for Bookstore Data*

3a. Gamma Distribution of Initial Visiting Rates



3b. Gamma Distribution of Updating Multiplier

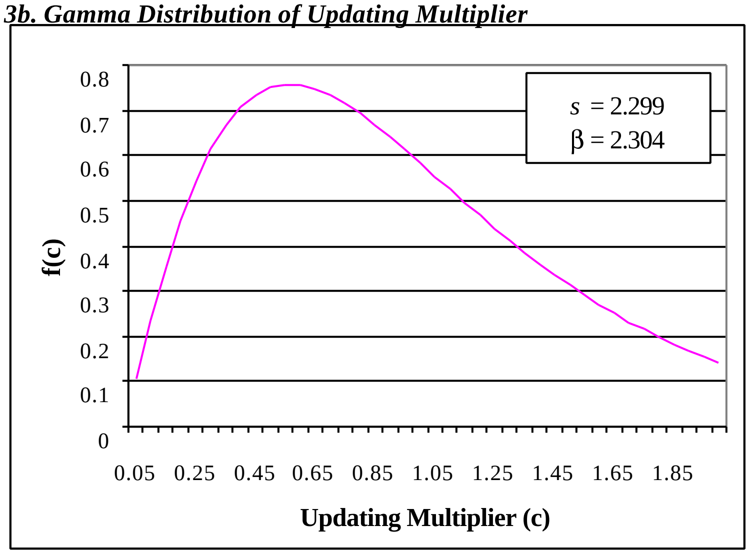


Figure 4. *Forecasts of Repeat Visits to the Bookstore*

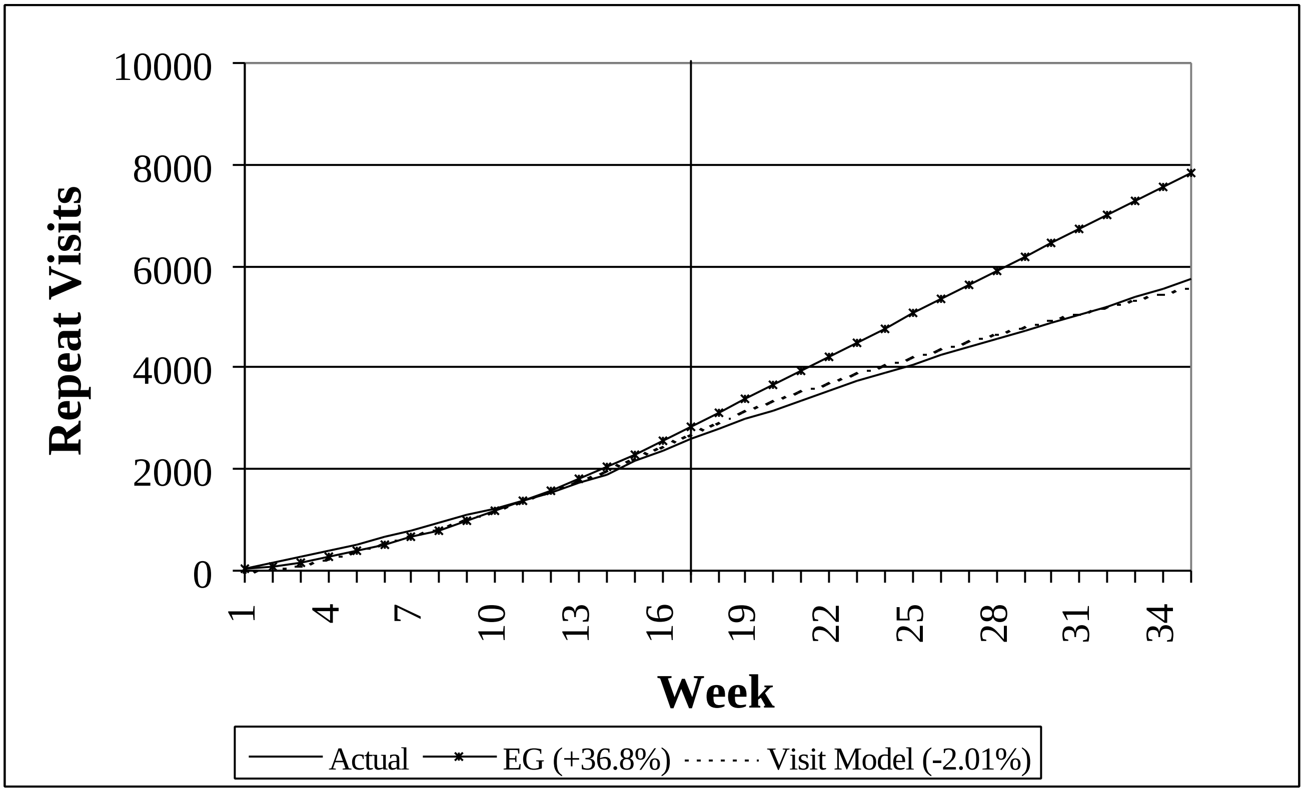


Table 1. *Summary of Visit Data over Time*

	BOOKSTORE		**CD STORE**	
	Months 1-4	**Months 5-8**	**Months 1-4**	**Months 5-8**
Total Number of Visits	5402	5899	1729	1890
Number of Unique Visitors	2693	2717	988	920
Visits / Visitor	2.01	2.17	1.75	2.05

Table 2. *Model Results for Bookstore*

	Stationary EG Model	**Evolving Visit Model**
r	0.483	0.324
a	42.955	16.857
s		2.299
b		2.304
LL	-34,347.2	-33,648.0
No. of parameters	2	4
CAIC	68,711.17	67,296.0

Table 3. *Model Results for CD Store*

	Stationary EG Model	**Evolving Visit Model**
r	0.255	0.165
a	28.305	8.889
s		2.084
b		2.104
LL	-9,459.6	-9,120.7
No. of parameters	2	4
CAIC	18,934.1	18,271.0

Table 4*. Bookstore Conversion Rates*

median=0.0349	**Decreasing Frequency**	**Increasing Frequency**
Infrequent Visitors	CELL 1 CR = 10.9% (N=526)	CELL 2 CR = 11.3%(N=510)
Frequent Visitors	CELL 3 CR = 14.6%(N=516)	CELL 4 CR = 18.6% (N=546)

Table 5. *CD Store Conversion Rates*

median=0.0431	**Decreasing Frequency**	**Increasing Frequency**
Infrequent Visitors	CELL 1 CR = 3.8% (N=129)	CELL 2 CR = 5.7% (N=161)
Frequent Visitors	CELL 3 CR = 4.0% (N=160)	CELL 4 CR = 7.6% (N=131)