

Modeling Browsing Behavior at Multiple Websites

Young-Hoon Park

Johnson Graduate School of Management, Cornell University, 330 Sage Hall, Ithaca,
New York 14853-6201, yp34@cornell.edu

Peter S. Fader

Wharton School, University of Pennsylvania, 749 Huntsman Hall, 3730 Walnut Street, Philadelphia,
Pennsylvania 19104-6340, faderp@wharton.upenn.edu

While there is a growing literature on investigating the Internet clickstream data collected for a single site, such datasets are inherently incomplete because they generally do not capture shopping behavior across multiple websites. A customer's visit patterns at one or more other sites may provide relevant information about the timing and frequency of his or her future visit patterns at the site of interest.

We develop a stochastic timing model of cross-site visit behavior to understand how to leverage information from one site to help explain customer behavior at another. To this end, we incorporate two sources of association in browsing patterns: one for the observable outcomes (i.e., arrival times) of two timing processes and the other for the latent visit propensities across a set of competing sites. This proposed multivariate timing mixture model can be viewed as a generalization of the univariate exponential-gamma model.

In our empirical analysis, we show that a failure to account for both sources of association not only leads to poor fit and forecasts, but also generates systematically biased parameter estimates. We highlight the model's ability to make accurate statements about the future behavior of the "zero class" (i.e., previous nonvisitors to a given site) using summary information (i.e., recency and frequency) from past visit patterns at a competing site.

Key words: Internet browsing behavior; data integration; multivariate duration models; customer acquisition

History: This paper was received May 27, 2002, and was with the authors 3 months for 3 revisions; processed by Scott A. Neslin.

1. Introduction

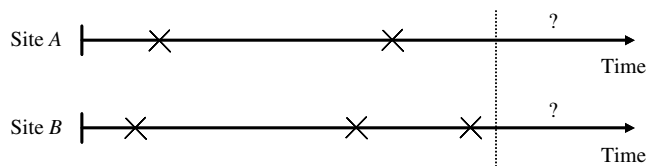
Virtually every commercial website monitors traffic to its own site and captures data on its own visitors. Given the observed behavior of these visitors, it is possible to develop a solid understanding of the customer base. This, in turn, can enable managers to predict likely future behavioral patterns and to determine customer response to future marketing efforts. As information technology continues to advance rapidly, the amount and richness of this type of clickstream data will continue to provide managers with a valuable tool to conduct their customer relationship management (CRM) efforts.

However, while there is a growing literature on investigating the Internet clickstream data collected for a single site (Bucklin and Sismeiro 2003, Johnson et al. 2003, Moe and Fader 2004a), such datasets are inherently incomplete because they generally do not capture shopping behavior across multiple websites. A customer's visit patterns at one or more other sites may provide relevant information about the timing and frequency of his or her future visit patterns at the site of interest.

Consider a hypothetical example of visiting behavior at two online retailers in a given product category, for example, books. Figure 1 illustrates the sequence of visits to site *A* and site *B* for a particular individual. Site *A* (e.g., Amazon.com) observes the upper series of visits over time and may want to predict when this customer will next visit the site and how often she will visit during a future period. Site *B* (e.g., arch-rival Barnesandnoble.com) may also want to make the same type of forecasting statements using the data it has available for this customer (lower series).

Using the kinds of models discussed in the papers cited above, managers could use each of these time series by itself to make separate predictions about future visit patterns at each site. But a key question is whether (and how) these predictions could be improved by taking into account the data from the other site as well. In this case, a clear pattern emerges from the combined data: Visits to site *B* are immediately followed by visits to site *A*. Thus, it is highly likely that this customer will visit site *A* right after the observation period ends (vertical dashed line)

Figure 1 Example of Internet Browsing Behavior



because she visited site *B* right before the end. Inferences based on the data from site *A* alone would be quite different from those drawn from the integrated database. We argue that there are a variety of cross-site patterns beyond the one shown in this stylized example, and that many of these patterns will be borne out in actual clickstream datasets.

The objective of this research is to better understand store visit patterns by developing a stochastic model that explicitly captures the critical features of cross-site timing behavior at the individual level. To this end, we consider two different types of cross-site dependence in store visit timing behavior: one for the observable outcomes (i.e., arrival times) of two timing processes and the other for the latent visit propensities across a set of competing sites. Based on the proposed approach, we can determine whether it is possible to make better inferences about individual-level browsing behavior at multiple sites using the combined database instead of assuming that visit patterns at different websites are independent of each other.

The importance of this issue is quite evident in the burgeoning practitioner literature on CRM. Many firms tout their capabilities to offer complete “360-degree” coverage of customer behavior, and industry experts emphasize the importance of “householding” (i.e., linking data from different sources to individual customers) in their data warehousing efforts (Swift 2001). While we focus on only one of the many possible benefits that arise from this type of data integration, the same modeling approach can be applied to other related domains, such as combining customer visit patterns across multiple channels (e.g., website, retail store) and combining activities across different business units for a given multiservice firm (e.g., banking, stock trading). The model presented here is sufficiently general to capture and describe these behaviors as well as it may work for cross-site visiting patterns.

Beyond the specification of the model per se, these different applications of data integration share several critical managerial issues that our model can help illuminate. Of particular interest is the desire to predict time-to-first-trial behavior at one site (or channel, etc.) given the pattern of past visits/usage at another. Prior researchers (Morrison and Schmittlein 1981) have observed that the so-called “zero class”

(i.e., the set of previous nonvisitors/buyers) from one period often accounts for more purchases in a subsequent period than any other single class of past customers. Our joint-timing model can successfully leverage the visit pattern from one site to allow us to make more informed statements about the number (and timing) of zero-class buyers who will make their initial visits to the other site sometime in the future.

The zero class merits explicit mention because it is the primary focus of customer acquisition efforts that have bedeviled online retailers in recent years. According to industry sources, such as the well-known Shop.org *State of Online Retailing* report (2003), e-commerce merchants are becoming increasingly concerned about their abilities to attract profitable new customers in an efficient manner. Our research provides a way to evaluate the “goodness” of customers at the individual level by capturing browsing patterns across online retailers.

Along these lines, we investigate relevant managerial issues related to the zero class. Imagine a given customer who has never visited a specific site (say, site *A*) but did visit a different site (site *B*) at least once in the past. How does his or her likelihood of visiting site *A* depend on the pattern of past visits to site *B*? Who would be a more valuable customer from site *A*’s perspective—someone who has visited site *B* fairly frequently (but not very recently) or someone who has visited site *B* fairly recently (but not very frequently)? Once again, it is easy to see that this type of question may be equally important for managers addressing issues about cross-channel and cross-business unit behavioral patterns as well as cross-site ones.

Finally, in addition to addressing these important substantive issues, this research also offers several novel methodological contributions. As noted earlier, our multivariate timing model, a generalization of the univariate exponential-gamma model, accounts for two different types of correlation in browsing patterns. We achieve this with closed-form solutions that enable us to use standard maximum likelihood parameter estimation techniques. The key to deriving this highly flexible density function is the use of the Sarmanov family of multivariate distributions (Kotz et al. 2000, Lee 1996). To the best of our knowledge, this is the first time this class of functions has been deployed in marketing, and we foresee many future applications for it across a wide variety of problem settings.

The remainder of the paper is organized as follows. Section 2 deals with the conceptual background that motivates our modeling approach. Section 3 gives an overview of the data and describes summary statistics

as a way to begin to understand some of the associations in browsing patterns that exist across competing websites. In §4, we present the model specification and use a simulation study to demonstrate the potential problems caused by ignoring these associations. In §5, we apply the proposed model to Internet clickstream data for websites in two different categories (books and music). Section 6 introduces conditional expectations for the model and examines the zero-class problem, that is, the likelihood and expectations for future visits by previous nonvisitors. We discuss other managerial implications of this research and conclude with directions for future research in §7.

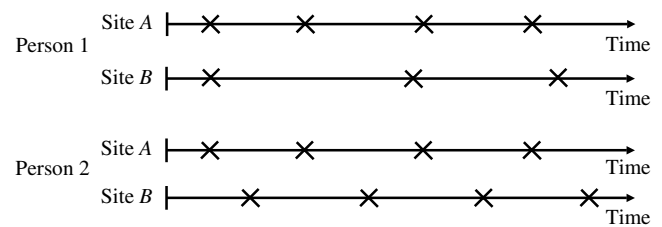
2. Conceptual Background

As discussed earlier, the vast majority of papers that discuss online visit behavior are strictly focused on the patterns that arise for a group of customers visiting a single website. This is consistent with the limited viewpoints provided by simple data-collection techniques, such as an examination of a site's log files. Padmanabhan et al. (2001) use the term *site-centric* to describe such a dataset and contrast it with a broader *user-centric* view that would also cover behavior for the same users at one or more other sites. Some researchers (e.g., Iyer and Pazgal 2003, Johnson et al. 2004) have taken steps to explicitly consider certain multiple-site aspects of online browsing (e.g., the number of different sites visited in a given month), but these papers do not attempt to understand or predict specific arrivals at each site.

Padmanabhan et al. (2001) conduct a number of analyses to demonstrate the significant biases that arise when analysts adopt a site-centric perspective, but they stop short of developing a new model that can bridge the gap between these two different views of customer behavior. We take on this challenge, starting with a model that assumes independence in cross-site behavior, then introducing two unique sources of association in browsing patterns: one for the observable outcomes (i.e., arrival times) of two timing processes and another for the unobservable visit propensities (regardless of the timing of the actual arrivals) across the set of competing sites. We illustrate these concepts with an example.

Figure 2 shows why it is important to distinguish between these two sources of association in cross-site browsing behavior. First, consider the browsing behavior of person 1 and person 2 only at site A. They appear to be identical in terms of the timing and frequency of visits at site A alone. If we only had site-centric data, we would likely conclude that their future behavior patterns at site A would be very similar. However, once we combine each person's browsing information at site B with the visits at site A,

Figure 2 Associations Across Browsing Patterns



we can begin to tell two very different stories about their shopping habits. When person 1 visits site A, she is likely to visit site B at nearly the same time. The underlying preference for both sites may not be similar, but she might be an avid comparison shopper, who chooses to look across alternative sites to ensure the best price or availability of an item. But the bottom line is that the association in *visit timing* is high for this customer.

In contrast, person 2 does not seem to have a high level of coincidence for his visits, but he does have a similar number of visits at both sites for the entire period of observation. Therefore, the similarity in *visit rates* may be high for him, but the coincidence of visit timing is not. He may visit different sites on the basis of his specific interests at each shopping occasion (e.g., browsing cookbooks at site A but visiting site B to find books for his children).

In general, misleading inferences may arise from ignoring either of these two associations; for example, suppose that a customer (like person 1 in Figure 2) has a high level of coincidence of visits across two sites. However, we do not take this tendency into account. Her desire to comparison shop will drive up the number of visits at one site to a level that exceeds her true, underlying desire/propensity to visit that site by itself. Thus, we might erroneously infer higher visit rates than her actual propensity at one or both sites. To the extent that some function of visit rates can serve as a valid indicator of customer value (Moe and Fader 2004a, b), we would then overestimate her potential worth as a customer. We will examine several biases like this one in the simulation study discussed in §4.

Therefore, to properly bridge the gap between the site-centric and user-centric perspectives, it is imperative that we allow for both sources of association. Before developing a model specification that accomplishes this goal, we first describe the datasets we will use for our empirical investigation. This will help motivate/clarify some of the analytic components of the model and will also provide us with some simple summary statistics to use as benchmarks to better understand and appreciate some of the model's parameters.

3. Data Overview

In this section, we describe the data used for this research and propose several sets of summary statistics derived from the combined database to investigate possible associations in cross-site visit behavior.

3.1. Data Description

We use Internet clickstream data collected by Media Metrix, Inc., which is now part of comScore Networks. At the time of our dataset, Media Metrix maintained a panel of approximately 20,000 panelists whose Internet behavior was recorded over time. These panelists had agreed to install special unobtrusive software on their computers that monitored their browsing activities. The collected data contain information regarding what sites individuals visit and when they visit. The data also include the precise day and time when panelists viewed a specific URL.

We are interested in the dates that each individual visits a given site. To consolidate the data, we aggregate panelist store visits at a given site to the daily level.¹ Any session in which the individual views a URL with the online store's domain name is considered a visit to that store. If a given individual visits a site multiple times in a single calendar day, that is coded as one visit on the day when the session began.

For our purposes, we use data pertaining to two major online retailers in each of two product categories: books (Amazon.com and Barnesandnoble.com) and music (CDNOW.com and Musicboulevard.com). The data span a period of eight months from October 1997 to May 1998. The total number of panelists who made at least one visit to either site in books and music was 4,955 and 2,422, respectively; the total number of visits at book and music sites was 12,640 and 5,387, respectively. We exclude people with no visits to both sites in each product category. The number of visits made by unique visitors at site *A*, site *B*, and both sites are detailed in Table 1, which shows that 24% (1,193 of 4,955 panelists) and 17% (422 of 2,422 panelists) of customers visited both sites in books and music, respectively.

The two online booksellers are quite different in terms of the number of unique visitors and visits from one another (i.e., Amazon.com dominates Barnesandnoble.com), but the two online stores in music are fairly similar to each other. These different

Table 1 Data Description for Books and Music

	Books		Music	
	Amazon.com	Barnesandnoble.com	CDNOW.com	Musicboulevard.com
Site <i>A</i>	Visitors	Visits	Visitors	Visits
Site <i>B</i>				
Site <i>A</i> only	2,681	5,143	927	1,625
Site <i>B</i> only	1,081	1,547	1,073	1,559
Both sites	1,193	5,950	422	2,203
Total	4,955	12,640	2,422	5,387

degrees of competitive parity will offer some interesting contrasts in our empirical analyses.

We further break down the number of visits at each site made by each individual in Tables 2 and 3 for books and music, respectively. As shown in these tables, the number of individuals who visited only one site accounts for a majority of panelists in both product categories: 76% (3,762 of 4,955 panelists) in books and 83% (2,000 of 2,422 panelists) in music. Since these customers did not visit both sites, they are a main source for online retailers in acquiring new customers.² These large groups are of particular interest in this research, and we will focus on them in §6.

3.2. Summary Statistics

We propose several different summary statistics to examine potential associations across sites. However, this is not an exhaustive set to infer possible dependence in the cross-site visiting process. While aggregate measures of summary statistics cannot disentangle underlying individual-level behavioral patterns in browsing data and are less valuable in the presence of sparse data, they may help detect the possible nature of associations before undertaking a model-based approach.

3.2.1. Coincidence of Visits. In this analysis, we concentrate on the customers who visited both sites to consider the degree of coincidence of visits. In Tables 4 and 5, we vary the length of time between visits to the two sites in books and music, respectively.

Table 4 can be read as follows: 364 panelists made 1106 same-day visits at both sites in books, which accounts for 18.59% of the total number of visits made by these 1193 customers and 8.75% of the total visits across the complete panel. In sum, nearly one-third of the visits that span both websites take place within the same week in books, and a similar (but slightly

¹ The data aggregation to the daily level may be questioned, but there are several solid justifications for this choice: (1) In the data used in this research, there are very few cases of multiple sessions at one site in a given day (less than 4% of all sessions at each of our focal sites); (2) methods for defining a session using Internet clickstream data are still somewhat controversial; and (3) other researchers (e.g., Moe and Fader 2004a, b) have successfully used data at the daily level.

² These presumed “zero-class” customers may have visited both sites at some point before our data period began, but we operate under the conservative assumption that our observed data reflect their tenure at each site. This corresponds with the course of action that practitioners would tend to use with similarly left-censored data.

Table 2 Number of Visits for Books

		No. of visits to Barnesandnoble.com					Total
		0	1	2	3	4+	
No. of visits to Amazon.com	0	—	812	178	47	44	1,081
	1	1,693	352	96	31	34	2,206
	2	500	166	62	26	27	781
	3	210	80	34	10	15	349
	4+	278	102	59	39	60	538
Total		2,681	1,512	429	153	180	4,955

lower) fraction characterizes the nature of coincidence in music (see Table 5). Figure 3 shows how these percentages of coincidental visits grow over a period of 30 days. It is interesting to see that the pattern of coincidental visits in books is very similar to that of music, despite the different competitive environments pointed out earlier.

Not only does this figure provide us with a useful initial understanding of the timing of two-site visit patterns, but it also will be the basis for a later validation test. Our model's ability to recover this time pattern will be a good way to showcase its capabilities and limitations.

3.2.2. Visit Rates. Unlike the above summary statistics based on the subset of customers who visited both sites, we now consider the entire set of customers in each category to look for possible patterns in the overall propensities to visit each site. We calculate the visit rate at each site as the number of visits divided by the observed time horizon for each individual (i.e., time since first observed visit to either site). To assess possible associations in visit propensities, we calculate the correlation between visit rates at both sites in each product category.

For the two booksellers this simple correlation is 0.0673, and for the two music retailers it is -0.0356 . These minimal correlations seem to suggest that there are no linked propensities in each category (in fact, the latter correlation indicates the presence of some site-specific inertia among music customers). But the proper way to estimate the visit rates for each site is to use a well-specified timing model that can capture the latent visit tendency for each person. Furthermore, by

Table 3 Number of Visits for Music

		No. of visits to Musicboulevard.com					Total
		0	1	2	3	4+	
No. of visits to CDNOW.com	0	—	818	163	44	48	1,073
	1	668	158	45	12	21	904
	2	131	49	16	8	13	217
	3	60	13	10	2	5	90
	4+	68	25	15	8	22	138
Total		927	1,063	249	74	109	2,422

Table 4 Coincidence of Visits for Books

Period	No. of people	No. of visits	% at both sites	% of all visits
Same day	364	1,106	18.59	8.75
2 days	414	1,321	22.20	10.45
3 days	451	1,460	24.54	11.55
4 days	491	1,581	26.57	12.51
5 days	529	1,704	28.64	13.48
6 days	556	1,808	30.39	14.30
Same week	583	1,909	32.08	15.10
Total	1,193	5,950		

building an explicit bivariate model, we will estimate a correlation parameter that specifically allows these rates to be linked.

4. Model Development

We build a model of timing behavior for databases combined across multiple websites. As a starting benchmark, we describe a simple bivariate model that assumes complete independence in cross-site visit timing behavior. We then address several issues that arise in developing a multivariate model; to accommodate them we introduce the Sarmanov family of multivariate distributions. This remarkable methodology provides some very desirable properties that assist in the specification, estimation, and interpretation of a multivariate model. We discuss how both forms of possible association (coincidence and linked propensities) are brought into the proposed model. We conclude the section with the results of a simulation that illustrates the possible biases caused by ignoring these forms of dependence in visit behavior across sites.

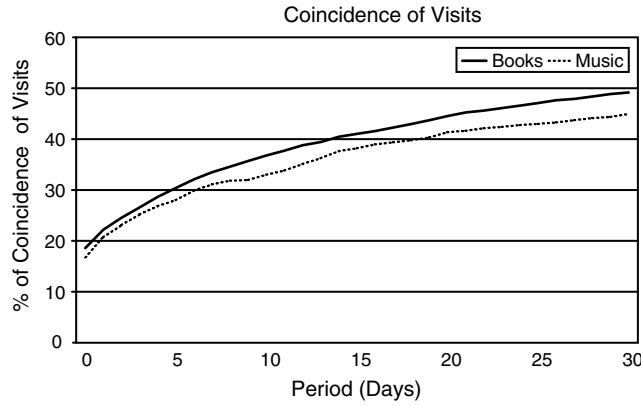
4.1. A Simple Bivariate Model

We start with a commonly used model that assumes complete independence in browsing patterns across sites. This section reviews the standard exponential-gamma mixture model, which has been applied extensively in the marketing literature because of its parsimony and performance. Each individual's intervisit times are assumed to be exponentially distributed, governed by λ_A^i for site A and λ_B^i for site B .

Table 5 Coincidence of Visits for Music

Period	No. of people	No. of visits	% at both sites	% of all visits
Same day	126	370	16.80	6.87
2 days	145	456	20.70	8.46
3 days	160	509	23.10	9.45
4 days	171	555	25.19	10.30
5 days	181	591	26.83	10.97
6 days	190	617	28.01	11.45
Same week	203	658	29.87	12.21
Total	422	2,203		

Figure 3 Coincidence of Visits



Naturally, some individuals have high visit rates, while others visit infrequently. This heterogeneity in visit rates is assumed to follow gamma distributions with parameters (r_A, α_A) for site A and (r_B, α_B) for site B . As stated, each site has completely different (independent) parameter sets. These distributions are given by the following densities:

$$f_s(t_{s_j}^i; \lambda_s^i) = \lambda_s^i e^{-\lambda_s^i(t_{s_j}^i - t_{s_{j-1}}^i)} \quad \text{and}$$

$$g_s(\lambda_s^i; r_s, \alpha_s) = \frac{\lambda_s^{r_s-1} \alpha_s^{r_s} e^{-\alpha_s \lambda_s^i}}{\Gamma(r_s)} \quad \text{for } s = A, B,$$

where λ_s^i is individual i 's latent rate of visit at site s , $t_{s_j}^i$ is the day when the j th repeat visit occurred at site s , and t_0^i is the day of her initial visit at either site. For a single visit at site s for individual i , this leads to the following exponential-gamma mixture model:

$$f_s(t_{s_j}^i; r_s, \alpha_s) = \int_0^\infty f_s(t_{s_j}^i; \lambda_s^i) \times g_s(\lambda_s^i; r_s, \alpha_s) d\lambda_s^i$$

$$= \frac{r_s}{\alpha_s} \left(\frac{\alpha_s}{\alpha_s + t_{s_j}^i - t_{s_{j-1}}^i} \right)^{r_s+1}.$$

For individual i 's single visit event at each site, therefore, a simple bivariate model ignoring any association in browsing patterns is the product of the two univariate exponential-gamma models:

$$f(t_{A_j}^i, t_{B_j}^i; r_A, \alpha_A, r_B, \alpha_B) = f_A(t_{A_j}^i; r_A, \alpha_A) \times f_B(t_{B_j}^i; r_B, \alpha_B)$$

$$= \frac{r_A}{\alpha_A} \left(\frac{\alpha_A}{\alpha_A + t_{A_j}^i - t_{A_{j-1}}^i} \right)^{r_A+1}$$

$$\cdot \frac{r_B}{\alpha_B} \left(\frac{\alpha_B}{\alpha_B + t_{B_j}^i - t_{B_{j-1}}^i} \right)^{r_B+1}. \quad (1)$$

A popular approach to obtain the likelihood function at site s for individual i is to specify the individual-level likelihood function at site s ,

conditional on that person's latent visit rate at that site. The likelihood at site s for individual i is the product of s_{ji} exponential terms, where s_{ji} is the number of repeat visits at site s made by the i th individual, times an additional term to account for the right censoring that occurs between that customer's last arrival at site s and the end of the observed calibration period (at time T):

$$[L_s^i; \lambda_s^i] = \lambda_s^i e^{-\lambda_s^i(t_{s_1}^i - t_0^i)} \cdot \lambda_s^i e^{-\lambda_s^i(t_{s_2}^i - t_{s_1}^i)} \dots \lambda_s^i e^{-\lambda_s^i(t_{s_{j_i}}^i - t_{s_{j_i-1}}^i)} \cdot e^{-\lambda_s^i(T - t_{s_{j_i}}^i)}. \quad (2)$$

To get the likelihood for the observed data (i.e., the unconditional distribution) at site s for individual i , we then integrate Equation (2) across all possible values of λ_s^i using the gamma mixing distribution:

$$[L_s^i; r_s, \alpha_s] = \int_0^\infty [L_s^i; \lambda_s^i] \times g_s(\lambda_s^i; r_s, \alpha_s) d\lambda_s^i$$

$$= \frac{\Gamma(r_s + s_{ji})}{\Gamma(r_s)} \left(\frac{\alpha_s}{\alpha_s + T - t_0^i} \right)^{r_s} \left(\frac{1}{\alpha_s + T - t_0^i} \right)^{s_{ji}}, \quad (3)$$

where $g_s(\lambda_s^i; r_s, \alpha_s)$ denotes the gamma distribution. Multiplication of the unconditional likelihoods at each site yields the simple bivariate model assuming complete independence in browsing patterns across websites, which can then be multiplied across the N individuals to get the overall likelihood:

$$L_{AB}^I = \prod_{i=1}^N \underbrace{\frac{\Gamma(r_A + A_{ji})}{\Gamma(r_A)} \left(\frac{\alpha_A}{\alpha_A + T - t_0^i} \right)^{r_A} \left(\frac{1}{\alpha_A + T - t_0^i} \right)^{A_{ji}}}_{\text{Unconditional likelihood at site A}}$$

$$\cdot \underbrace{\frac{\Gamma(r_B + B_{ji})}{\Gamma(r_B)} \left(\frac{\alpha_B}{\alpha_B + T - t_0^i} \right)^{r_B} \left(\frac{1}{\alpha_B + T - t_0^i} \right)^{B_{ji}}}_{\text{Unconditional likelihood at site B}}. \quad (4)$$

As shown in Equation (4), the independent model requires only the information on the observed calibration period ($T - t_0^i$) and the frequency of repeat visits at each site (A_{ji} and B_{ji}) at the individual level. It implies that these three elements represent sufficient information to estimate all four parameters of the independent model.

In other words, the independent bivariate model ignores most of the information of visit timing at each site. As long as the data required to estimate the model (i.e., $T - t_0^i$, A_{ji} , and B_{ji}) are the same, this model will yield exactly the same parameter values regardless of the timing of the actual arrivals. This result is highly counterintuitive and suggests some doubt about the theoretical basis (and empirical performance) of the independent model. Indeed, the

correlated bivariate model does not share these simplistic data requirements. We now turn to our consideration of a more rigorous bivariate model that will use all of the timing data available from both sites while allowing for formal associations in the timing patterns across the two sites.

4.2. Issues in Developing a Multivariate Model

The most common way to introduce dependence is to begin with two independent univariate distributions and combine them using a third multiplicative term that ties together the two random variables of interest. For instance, Farlie (1960) proposed a general form of bivariate distribution functions:

$$F(t_A, t_B) = F_A(t_A) \times F_B(t_B) \times \{1 + \alpha \varphi_A(t_A) \varphi_B(t_B)\},$$

where the function $F(t_A, t_B)$ is the joint cumulative distribution, the functions $F_A(t_A)$ and $F_B(t_B)$ are the marginal cumulative distributions, and α is a parameter measuring association. It suffices if we choose for $\varphi_A(t_A)$ and $\varphi_B(t_B)$ functions that are bounded and have bounded first derivatives with respect to their arguments.

Given this setup, the function $F(t_A, t_B)$ of the Farlie-Gumbel-Morgenstern (FGM) family, which has been studied extensively in bivariate model building (e.g., Johnson and Kotz 1975, 1977), is given by

$$F(t_A, t_B) = F_A(t_A) \times F_B(t_B) \cdot \{1 + \alpha[1 - F_A(t_A)][1 - F_B(t_B)]\}, \quad |\alpha| \leq 1. \quad (5)$$

If the densities $f(\cdot)$ corresponding to $F(\cdot)$ exist, then Equation (5) implies

$$\begin{aligned} f(t_A, t_B) &= \frac{\partial^2 F(t_A, t_B)}{\partial t_A \partial t_B} \\ &= f_A(t_A) \times f_B(t_B) \\ &\quad \cdot \{1 + \alpha[1 - 2F_A(t_A)][1 - 2F_B(t_B)]\}. \end{aligned}$$

Chintagunta and Haldrar (1998) appear to be the first to use the FGM approach in marketing as a model of purchase timing in related categories (e.g., pasta and pasta sauce). While FGM offers a reasonably straightforward way to introduce dependence, this approach is very limited from an inferential standpoint. In particular, FGM does *not* generally yield marginal distributions that match the functional forms of the designated univariate densities.

In other words, when one random variable is isolated by integrating over all possible values of the other random variable, it is highly desirable to get resulting marginal densities that are identical to the univariate densities of interest: $f_A(t_A) = \int_{-\infty}^{\infty} f(t_A, t_B) dt_B$ and $f_B(t_B) = \int_{-\infty}^{\infty} f(t_A, t_B) dt_A$. In the

case of the bivariate exponential distribution, the FGM family is fine in this regard—it yields the standard exponential as its marginal densities. However, for most other density functions, for example, the gamma distribution, this property does not hold for FGM. So while it is possible to formulate a bivariate gamma distribution using FGM, the resulting distribution would not be easy to interpret or manipulate.

In the next section, we introduce a relatively unknown family of multivariate distributions that overcomes this critical shortcoming and offers some other useful properties as well.

4.3. The Sarmanov Family of Bivariate Distributions

The model formulation described in this research is based on the Sarmanov family of bivariate distributions (Kotz et al. 2000, Lee 1996). Sarmanov (1966) first described a highly versatile family of bivariate densities, but his work went largely unnoticed until Lee (1996) introduced it to the mainstream statistics literature.

The fundamental idea in constructing the Sarmanov family of bivariate distributions is very similar to that of other bivariate distributions, such as FGM. The main difference lies in how dependence is brought into the bivariate distribution. The Sarmanov family of bivariate densities is highly flexible and includes many of the FGM distributions as special cases.

Assume that $f_A(t_A)$ and $f_B(t_B)$ are univariate probability density functions. Let $\phi_A(t_A)$ and $\phi_B(t_B)$ be bounded nonconstant mixing functions such that $\int_{-\infty}^{\infty} \phi_s(t_s) f_s(t_s) dt_s = 0$ for $s = A, B$. The function $f(t_A, t_B)$ of the Sarmanov family of bivariate distributions is defined by

$$f(t_A, t_B) = f_A(t_A) \times f_B(t_B) \times \{1 + \zeta \phi_A(t_A) \phi_B(t_B)\}. \quad (6)$$

This is a bivariate joint density with specified marginals $f_A(t_A)$ and $f_B(t_B)$, provided ζ is a real number that satisfies the condition $1 + \zeta \phi_A(t_A) \phi_B(t_B) \geq 0$ for all t_A and t_B . ζ is interpreted as an unnormalized correlation (i.e., covariance), and the correlation coefficient ρ_ζ of t_A and t_B , if it exists, is given by

$$\rho_\zeta = \text{Corr}[t_A, t_B] = \frac{\zeta \nu_A \nu_B}{\sigma_A \sigma_B},$$

where $\nu_s = \int_{-\infty}^{\infty} t_s \phi_s(t_s) f_s(t_s) dt_s$, $\sigma_s^2 = \int_{-\infty}^{\infty} (t_s - \mu_s)^2 \cdot f_s(t_s) dt_s$, and $\mu_s = \int_{-\infty}^{\infty} t_s f_s(t_s) dt_s$ for $s = A, B$. The correlation coefficient ρ_ζ of t_A and t_B is bounded by

$$|\rho_\zeta| \leq |\zeta| \sqrt{E[\phi_A^2(t_A)] E[\phi_B^2(t_B)]}.$$

Lee (1996) discusses methods to find the mixing functions $\phi_s(t_s)$ for distributions of the multivariate natural exponential families. Assume that

$f_s(t_s)$ is defined on $[0, \infty)$ for $s = A, B$. Let $L_s(k) = \int_0^\infty e^{-kt_s} f_s(t_s) dt_s$ denote the Laplace transform of $f_s(t_s)$. Define $\phi_s(t_s) = e^{-t_s} - L_s(1)$ for $t_s \geq 0$. Then the function $f(t_A, t_B)$ defined in Equation (6) is a bivariate density with designated marginals $f_s(t_s)$ for $s = A, B$. We later derive the Sarmanov family of bivariate exponential and gamma densities on the basis of this method (see Lee 1996 for other general methods to find $\phi_s(t_s)$).

It is of interest to investigate the properties of the Sarmanov family of multivariate distributions. In contrast to FGM and other bivariate families, marginals of this class of distributions are guaranteed to take on the desired univariate densities. The Sarmanov family is the only class of multivariate distributions that offers this highly desirable property.

A secondary benefit of the Sarmanov family is that the range of its correlation coefficients is generally wider than that of FGM distributions. The range shown above can still be fairly restrictive in many cases; it does not generally allow for the full spectrum of $(-1, 1)$, but at least the Sarmanov family offers more flexibility without any loss in parsimony or analytical convenience.

In the next section we develop a bivariate timing mixture model, which is the first use of the Sarmanov family in such a context. This is a very practical application that takes great advantage of the Sarmanov properties in order to address the issues discussed throughout the paper. We will use the Sarmanov approach in two different ways simultaneously to capture two sources of association: one for the similarity in arrival times of two timing processes (via a Sarmanov bivariate exponential distribution) and another for the similarity in latent visit propensities across a pair of competing sites (via a Sarmanov bivariate gamma distribution).

4.4. The Proposed Model

Similar to a standard timing mixture model such as the aforementioned exponential-gamma, the proposed model has two main components: an individual-level timing process and a cross-individual heterogeneity distribution, each of which is constructed using the Sarmanov approach.

4.4.1. Timing Process. We begin by developing the correlated bivariate exponential distribution for cross-site visit-timing processes, with a univariate exponential timing model marginally at each site.

Assume that individual i 's latent visit rates are λ_A^i and λ_B^i for site A and site B , respectively. The Laplace transform of $f_s(t_s; \lambda_s^i)$ has the form

$$L_s^f(k; \lambda_s^i) = \frac{\lambda_s^i}{k + \lambda_s^i}$$

for $s = A, B$. Let $\phi_s(t_{s_j}^i - t_{s_{j-1}}^i; \lambda_s^i) = e^{-(t_{s_j}^i - t_{s_{j-1}}^i)} - L_s^f(1)$ denote the required Sarmanov mixing functions. Using $\phi_s(t_{s_j}^i - t_{s_{j-1}}^i; \lambda_s^i) = e^{-(t_{s_j}^i - t_{s_{j-1}}^i)} - \lambda_s^i / (1 + \lambda_s^i)$, we can construct a bivariate exponential density:

$$\begin{aligned} f(t_{A_j}^i, t_{B_j}^i; \lambda_A^i, \lambda_B^i, \tau) \\ = f_A(t_{A_j}^i; \lambda_A^i) \times f_B(t_{B_j}^i; \lambda_B^i) \\ \cdot \{1 + \tau \phi_A(t_{A_j}^i - t_{A_{j-1}}^i; \lambda_A^i) \phi_B(t_{B_j}^i - t_{B_{j-1}}^i; \lambda_B^i)\}. \end{aligned} \quad (7)$$

This is a bivariate exponential density with exponential marginals $f_A(t_{A_j}^i; \lambda_A^i)$ and $f_B(t_{B_j}^i; \lambda_B^i)$. The correlation coefficient, ρ_τ , which captures the dependence in coincidence of visits (i.e., similarity in visit times), is given by

$$\rho_\tau = \tau \frac{\lambda_A^i}{(1 + \lambda_A^i)^2} \frac{\lambda_B^i}{(1 + \lambda_B^i)^2}, \quad (8)$$

where τ is a real number on the range

$$\begin{aligned} & \frac{-1}{\max\{L_A^f(1)L_B^f(1), (1 - L_A^f(1))(1 - L_B^f(1))\}} \\ & \leq \tau \leq \frac{1}{\max\{L_A^f(1)(1 - L_B^f(1)), (1 - L_A^f(1))L_B^f(1)\}}. \end{aligned}$$

Although this bivariate exponential model is a clear generalization of the univariate exponential timing model, there are certain familiar properties of the univariate exponential timing model that are not generalized in the multivariate case. These include the notion of a constant hazard rate as well as the exact interplay between the (univariate) exponential timing model and the Poisson counting model. Unlike the constant hazard rate in the univariate exponential timing model, the hazard rate of the bivariate exponential timing distribution is not constant in the presence of dependence, because the hazard rate at a given site (say, site A) is a function of intervisit times at the other site as well as past arrival times at site A . Therefore, it is necessary to account for the elapsed time since the last visit at each site in constructing the correlated timing process across websites. Appendix A describes the discussion of hazard functions for this model.

In addition, while translating the univariate exponential timing model into the Poisson counting model is straightforward, it is not analytically feasible to derive an equivalent counting model from the timing model in the bivariate context. Thus, simulations are required to generate the counts from the bivariate timing model. Alternatively, one could develop a Sarmanov version of the Poisson counting model, but it would not lend itself to a closed-form interarrival timing process.

Finally, it may be tempting to draw inferences directly from Equation (8), but it is not easy to do so in

a meaningful manner. First of all, in the next section, we bring in heterogeneity in the rate parameters, which will make it much harder to visualize λ_A^i and λ_B^i . Second, and more important, for mathematical convenience we will assume homogeneity in ρ_τ . Therefore, in the mixture model we actually estimate, there is not a perfect one-to-one mapping between the rate parameters and the correlation term.³

4.4.2. Heterogeneity Distribution. To capture heterogeneous latent visit propensities at two sites across the population, we employ the Sarmanov approach once again, but this time using the gamma distribution in order to obtain a bivariate gamma distribution with gamma marginals. The Laplace transform of $g_s(\lambda_s^i)$ has the form $L_s^g(k; r_s, \alpha_s) = (\alpha_s/(k + \alpha_s))^{r_s}$ for $s = A, B$. Let $\psi_s(\lambda_s^i; r_s, \alpha_s) = e^{-\lambda_s^i} - L_s^g(1)$ define the mixing functions. Using $\psi_s(\lambda_s^i; r_s, \alpha_s) = e^{-\lambda_s^i} - (\alpha_s/(1 + \alpha_s))^{r_s}$, we can construct a bivariate gamma density:

$$\begin{aligned} g(\lambda_A^i, \lambda_B^i; r_A, \alpha_A, r_B, \alpha_B, \omega) \\ = g_A(\lambda_A^i; r_A, \alpha_A) \times g_B(\lambda_B^i; r_B, \alpha_B) \\ \cdot \{1 + \omega \psi_A(\lambda_A^i; r_A, \alpha_A) \times \psi_B(\lambda_B^i; r_B, \alpha_B)\}. \end{aligned} \quad (9)$$

This is a bivariate gamma density with gamma marginals $g_A(\lambda_A^i; r_A, \alpha_A)$ and $g_B(\lambda_B^i; r_B, \alpha_B)$. The correlation coefficient, ρ_ω , which captures the dependence in underlying visit propensities (i.e., similarity in latent visit rates), is given by

$$\rho_\omega = \omega \frac{\sqrt{r_A}}{1 + \alpha_A} \left(\frac{\alpha_A}{1 + \alpha_A} \right)^{r_A} \frac{\sqrt{r_B}}{1 + \alpha_B} \left(\frac{\alpha_B}{1 + \alpha_B} \right)^{r_B}, \quad (10)$$

where ω is a real number bounded by

$$\begin{aligned} & \frac{-1}{\max\{L_A^g(1)L_B^g(1), (1 - L_A^g(1))(1 - L_B^g(1))\}} \\ & \leq \omega \leq \frac{1}{\max\{L_A^g(1)(1 - L_B^g(1)), (1 - L_A^g(1))L_B^g(1)\}}. \end{aligned}$$

A very useful property of a bivariate distribution developed using the Sarmanov method is that it can be expressed as a linear combination of products of the univariate densities. For instance, the Sarmanov bivariate gamma distribution with gamma marginals can be expressed as follows:

$$g(\lambda_A^i, \lambda_B^i; r_A, \alpha_A, r_B, \alpha_B, \omega)$$

³ We examined a variety of finite mixture models as alternatives to our parametric specification. One major advantage of this approach is that it allows for heterogeneity in the arrival correlations; that is, customers can have differing degrees of coincidence. But these finite mixture models perform far worse than the proposed model without offering any gains in parsimony or managerial diagnostics.

$$\begin{aligned} & = g_A(\lambda_A^i; r_A, \alpha_A) \times g_B(\lambda_B^i; r_B, \alpha_B) + \omega L_A^g(1) L_B^g(1) \\ & \cdot \left[\begin{aligned} & g_A(\lambda_A^i; r_A, \alpha_A + 1) \times g_B(\lambda_B^i; r_B, \alpha_B + 1) \\ & - g_A(\lambda_A^i; r_A, \alpha_A + 1) \times g_B(\lambda_B^i; r_B, \alpha_B) \\ & - g_A(\lambda_A^i; r_A, \alpha_A) \times g_B(\lambda_B^i; r_B, \alpha_B + 1) \\ & + g_A(\lambda_A^i; r_A, \alpha_A) \times g_B(\lambda_B^i; r_B, \alpha_B) \end{aligned} \right]. \end{aligned}$$

Although this expression appears to be much more complex than Equation (9), this decomposition facilitates the integration of the exponential with the gamma distribution and greatly reduces the complexity of the posterior distribution that emerges as a result. We address this integration to derive the unconditional timing model next.

4.4.3. The Complete Model. We derive a multivariate timing mixture model with closed-form analytic expressions. To get the unconditional distribution, we need to integrate Equation (7), the conditional timing process, across all possible values of λ_A^i and λ_B^i using Equation (9), the bivariate gamma density.⁴ As an illustrative example, the unconditional distribution for a single visit at each site made by individual i is as follows:

$$\begin{aligned} & f(t_{A_j}^i, t_{B_j}^i) \\ & = \int_0^\infty \int_0^\infty f(t_{A_j}^i, t_{B_j}^i; \lambda_A^i, \lambda_B^i, \rho_\tau) \\ & \quad \cdot g(\lambda_A^i, \lambda_B^i; r_A, \alpha_A, r_B, \alpha_B, \omega) d\lambda_A^i d\lambda_B^i \\ & = f_A(t_{A_j}^i; r_A, \alpha_A) \times f_B(t_{B_j}^i; r_B, \alpha_B) \\ & \quad \cdot \left[\begin{aligned} & 1 + \rho_\tau h_{A_j}^1(t_{A_j}^i; r_A, \alpha_A) \times h_{B_j}^1(t_{B_j}^i; r_B, \alpha_B) \\ & + \omega h_{A_j}^2(t_{A_j}^i; r_A, \alpha_A) \times h_{B_j}^2(t_{B_j}^i; r_B, \alpha_B) \\ & + \rho_\tau \omega h_{A_j}^3(t_{A_j}^i; r_A, \alpha_A) \times h_{B_j}^3(t_{B_j}^i; r_B, \alpha_B) \end{aligned} \right], \end{aligned} \quad (11)$$

where

$$\begin{aligned} h_{s_j}^1(t_{s_j}^i; r_s, \alpha_s) & = \left\{ e^{-(t_{s_j}^i - t_{s_{j-1}}^i)} \left(2 + \frac{\alpha_s^O}{r_s} + \frac{r_s^O}{\alpha_s^O} \right) - \left(1 + \frac{r_s^O}{\alpha_s^O} \right) \right\}, \\ h_{s_j}^2(t_{s_j}^i; r_s, \alpha_s) & = \left\{ \left(\frac{\alpha_s^O}{1 + \alpha_s^O} \right)^{r_s^O} - \left(\frac{\alpha_s}{1 + \alpha_s} \right)^{r_s} \right\}, \end{aligned}$$

⁴ To derive the unconditional distribution with closed-form expressions, it is necessary to use ρ_τ , not τ , in the bivariate exponential distribution. However, we can use either ω or ρ_ω for the bivariate gamma density in deriving the closed-form expression. For the sake of convenience, we use ω , but translating the resulting parameter estimates to obtain ρ_ω is straightforward.

$$h_{s_j}^3(t_{s_j}^i; r_s, \alpha_s) = \left\{ \left(\frac{\alpha_s^O}{1 + \alpha_s^O} \right)^{r_s^O} \left[e^{-(t_{s_j}^i - t_{s_{j-1}}^i)} \left(2 + \frac{1 + \alpha_s^O}{r_s} + \frac{r_s^O}{1 + \alpha_s^O} \right) - \left(1 + \frac{r_s^O}{1 + \alpha_s^O} \right) \right] - \left(\frac{\alpha_s}{1 + \alpha_s} \right)^{r_s} \cdot \left[e^{-(t_{s_j}^i - t_{s_{j-1}}^i)} \left(2 + \frac{\alpha_s^O}{r_s} + \frac{r_s^O}{\alpha_s^O} \right) - \left(1 + \frac{r_s^O}{\alpha_s^O} \right) \right] \right\},$$

and $r_s^O = r_s + 1$ and $\alpha_s^O = \alpha_s + t_{s_j}^i - t_{s_{j-1}}^i$ for $s = A, B$.

4.4.4. Constructing the Likelihood Function.

While the “housekeeping” for this expression requires some care, the actual mathematics are quite simple and the computational demands are negligible. Even when examining multiple purchases across many panelists, the amount of computation required to evaluate the complete likelihood function (see Appendix B) is on a par with that of the independent model discussed earlier. As seen in the relationship between Equation (11) and Equation (1), the likelihood function of the complete model (Equation (24) in Appendix B) collapses down to Equation (4) when both forms of cross-site dependence are ignored.

We leave the detailed discussion of the complete likelihood function to Appendix B, but here we briefly describe the steps required to construct it.

As noted earlier, the usual “memoryless” characteristic of an exponential model no longer applies in the bivariate case. Therefore, we must categorize each observation by the site currently visited and previously visited. (In this sense, the model closely resembles a Markov process, with the added benefit of capturing the time between transitions.) There are three possible current observations: (1) visit to site A, (2) visit to site B, and (3) no visit at either site because of right censoring (end of model calibration period). Likewise, there are three possible past observations: (1) visit to site A, (2) visit to site B, and (3) no visit at either site because of left censoring (start of model calibration period). Hence, there are nine different types of visits that we need to account for.

Each of these nine cases will generate a different probability expression that reflects the basic structure of Equation (11) but captures the unique timing elements required for that particular case. Appendix B derives the specific form for each of these nine expressions, starting at the individual level, bringing in the bivariate gamma mixing distribution, and finally yielding the likelihood function for the complete model (with nine distinct terms) in Equation (24).

A natural question to ask at this point is whether/how this bivariate model can be extended to allow for visit patterns at three or more sites. Such an extension is surprisingly simple, once the analyst is comfortable with this idea of constructing a separate likelihood

expression for each of the nine cases. For example, a trivariate model would replace this 3×3 setup with a 4×4 grid to capture all the possible transition patterns among three sites, and the specification for each of these 16 cases would still be consistent with Equation (11). More details about the derivation and operational aspects of the general multivariate model can be obtained in a technical appendix that is available from the authors.

4.5. Simulations

At the heart of this model lie two different types of cross-site associations: one that captures coincidence in the observed arrival times at each site and one that captures similarities (or dissimilarities) in the underlying rates that govern the visiting patterns at each site. A failure to account for either or both forms of association may lead to a misrepresentation of the true nature of browsing behavior within and across the two sites. To illustrate this point more clearly, we present results from simulations in which the true browsing processes across two sites are preset using different interdependent structures. The main purpose of these simulations is to demonstrate the nature and magnitude of the potential problems (i.e., biases) caused by ignoring one or both of the associations in browsing patterns.

We created different simulated datasets based on all six parameters: (1) two gamma parameters at each site: (r_A, α_A) for site A and (r_B, α_B) for site B; (2) the association in cross-site timing behavior across websites, ρ_τ ; and (3) the association in latent visit rates across the population, ρ_ω .

The first step is to generate an individual’s latent visit rates at both sites. We draw individual i ’s latent visit propensity at one site (say, site B) from the marginal gamma density, $g_B(\lambda_B^i; r_B, \alpha_B)$ of the Sarmanov bivariate gamma distribution, and then generate her underlying visit rate at the other site from the conditional gamma density, $g_A(\lambda_A^i | \lambda_B^i; r_A, \alpha_A, \rho_\omega)$.

The next step is to simulate intervisit times at both sites for the entire period of simulation. To do so, we draw an intervisit time based on the joint exponential timing model and probabilistically determine which site the individual would visit. This process of drawing intervisit times is repeated until the sum of intervisit times goes beyond the length of the simulation period.

For the gamma parameters $(r_A, \alpha_A, r_B, \alpha_B)$, we use the average visit rate for each site (r_A/α_A and r_B/α_B) and the coefficient of variation ($1/\sqrt{r_A}$ and $1/\sqrt{r_B}$) instead of using the gamma parameters directly. Given a set of the gamma parameters, we select three levels (i.e., minimum value, 0, and maximum value)

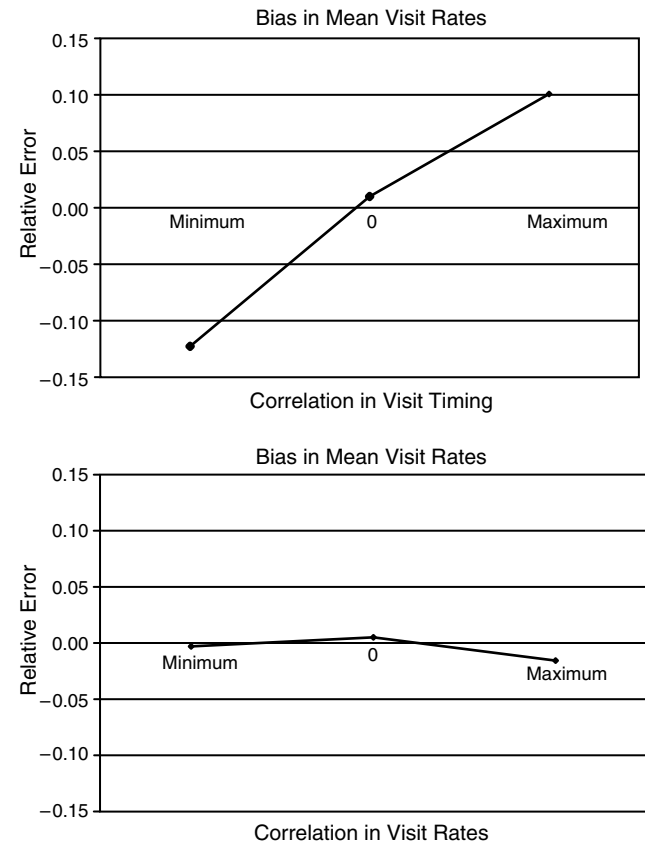
Table 6 Simulation Design

Profile	r_A/α_A	$1/\sqrt{r_A}$	r_B/α_B	$1/\sqrt{r_B}$	ρ_τ	ρ_ω
1	0.50	0.50	0.50	0.50	Minimum	Minimum
2	0.50	1.00	1.00	1.50	0.00	0.00
3	0.50	1.50	1.50	1.00	Maximum	Maximum
4	1.00	0.50	1.00	1.00	0.00	Maximum
5	1.00	1.00	1.50	0.50	Maximum	Minimum
6	1.00	1.50	0.50	1.50	Minimum	0.00
7	1.50	0.50	1.50	1.50	0.00	Minimum
8	1.50	1.00	0.50	1.00	Maximum	0.00
9	1.50	1.50	1.00	0.50	Minimum	Maximum
10	0.50	0.50	1.50	1.00	Minimum	0.00
11	0.50	1.00	0.50	0.50	0.00	Maximum
12	0.50	1.50	1.00	1.50	Maximum	Minimum
13	1.00	0.50	0.50	1.50	Maximum	Maximum
14	1.00	1.00	1.00	1.00	Minimum	Minimum
15	1.00	1.50	1.50	0.50	0.00	0.00
16	1.50	0.50	1.00	0.50	Maximum	0.00
17	1.50	1.00	1.50	1.50	Minimum	Maximum
18	1.50	1.50	0.50	1.00	0.00	Minimum

for ρ_τ and ρ_ω .⁵ We use a fractional factorial design consisting of 18 profiles, which are described in Table 6. We simulate 30 different datasets for each of the 18 profiles and then estimate the independent model, Equation (4), to determine the effects of ignoring the correlations.

For our error measure we use the relative difference between the estimated mean visit rate and the true underlying mean rate (i.e., (Estimated – Actual)/Actual). If this difference tends to be small, we have evidence that the independent model is robust to misspecifications involving one or both correlation terms. We use this relative error measure as a dependent variable in an ordinary least-squares regression, using the design matrix (i.e., Table 6) as explanatory variables in order to learn how the various model parameters may lead to possible biases. We use effects coding for the design matrix and run a single regression across all 540 simulated conditions (18 profiles \times 30 simulations per profile). Standard regression diagnostics did not reveal any problems with this methodological approach.

The results of this regression reveal the importance of accounting for coincidence in cross-site visit patterns. The top panel in Figure 4 illustrates that errors in our ability to capture the mean are directly related

Figure 4 Simulation Results of Interdependent Browsing Patterns

to the value of the correlation term for coincident visit timing, ρ_τ . As discussed in §2, this result makes intuitive sense: A behavioral tendency towards coincidence will lead to an exaggerated estimate of the visit rate at one or both sites. This effect is highly significant ($F_{2,527} = 197.22$, $p < 0.0001$). On the other hand, if this correlation term were negative (e.g., if visits to one site slowed down arrivals at the other site), then we would see a negative bias or an understatement of the true visit tendencies.

In sharp contrast to the effects for coincidence, the bottom panel in Figure 4 shows that there is a much smaller bias when the correlation in the underlying visit propensities, ρ_ω , is ignored. This effect is still significant from a purely statistical perspective ($F_{2,527} = 7.35$, $p < 0.01$), but it is not associated with a clear bias, as in the upper panel. Basically, ignoring any correlation in visit propensities (positive or negative) leads to a slight downward bias in the estimated average visit rate.

The difference in the magnitude of these bias estimates seems to suggest that a failure to account for coincidence in visits is more critical than accommodating correlation in latent visit propensities. Of course, the presence (or absence) of bias in these estimated means is only one indication of the overall

⁵ Given a set of the gamma parameters, we can directly set bounds for ρ_ω , since this correlation is a function of the gamma parameters, as described in Equation (10). On the other hand, ρ_τ is a function of latent visit propensities, λ_A^i and λ_B^i . We therefore need to integrate Equation (8) across all possible values of λ_A^i and λ_B^i to set the range of ρ_τ :

$$\begin{aligned} & -1 \\ & \max\{(1 + r_A/\alpha_A)(1 + r_B/\alpha_B), (1 + \alpha_A/r_A)(1 + \alpha_B/r_B)\} \\ & \leq \rho_\tau \leq \frac{1}{\max\{(1 + r_A/\alpha_A)(1 + \alpha_B/r_B), (1 + \alpha_A/r_A)(1 + r_B/\alpha_B)\}}. \end{aligned}$$

Table 7 Overview of Models

Models	Correlated timing	Correlated rates	No. of parameters
Model 1	—	—	4
Model 2	✓	—	5
Model 3	—	✓	5
Model 4	✓	✓	6

importance of accounting for (or ignoring) these correlations. A more telling diagnostic is to examine differences in model fit that arise from various nested models estimated on actual visit data. This is what we address in the next section.

5. Empirical Applications

To conduct our empirical analyses, we first created a systematic sample of the Media Matrix clickstream data based on the number of visits to site *A* (Amazon.com in books and CDNOW.com in music) and the number of visits to site *B* (Barnesandnoble.com in books and Musicboulevard.com in music). Specifically, we sorted the data based on the number of visits at site *A* and the number of visits at site *B* as the primary and secondary dimensions; then we selected every tenth individual for books and every fifth panelist for the music category. This resulted in samples of 493 and 482 individuals for books and music, respectively.

The models we examine in our empirical analysis vary on two dimensions: whether the model allows for correlation in visit timing (coincidence) and whether the model allows for correlation in the latent visit propensities. Hence, there are four different models (see Table 7), ranging from the independent model, Equation (1), to the complete model that accounts for both associations, Equation (11). For instance, model 2 in Table 7 uses a Sarmanov bivariate exponential timing model with two independent univariate gamma mixing distributions to derive a five-parameter multivariate timing mixture model.⁶

In addition, we estimate a model based on the FGM bivariate exponential distribution, which, like the Sarmanov model, yields the standard exponential as its marginal densities. In order to capture heterogeneous latent visit propensities at two sites, we employ two independent gamma densities. The resulting model is directly comparable to our model 2.⁷

⁶ We examined other specifications for the independent model, for example, the expo-power distribution (Saha and Hilton 1997, Seetharaman and Chintagunta 2003). This model fits worse than any of the exponential-gamma models. One reason for this result is that the expo-power model—despite the fact that it has three parameters—does not easily allow for unobserved heterogeneity across the population.

⁷ To build the complete likelihood function for the FGM specification, we follow the same procedure used to derive the

All models were estimated using standard constrained optimization code in the MATLAB programming language. Each model required less than one minute to run on a standard desktop computer. No irregularities or estimation problems (e.g., local optima or sensitivity to starting conditions) were observed. Despite the complexity required to derive the likelihood function, these models are very well behaved and efficient when it comes to parameter estimation.

5.1. Model Results

Tables 8 and 9 show the parameter estimates for books and music, respectively. It comes as no surprise that the complete model (model 4) provides the best fit while the independent model (model 1) is worst in both categories. Of more interest is the relative performance of the two intermediate models, which each capture only one of the correlation terms. In both categories, model 2, which accounts only for the correlation in coincidence of visit timing, is superior to model 3, which assumes independent timing but correlated rates. This is consistent with the simulation result showing that accounting for coincidence in visits is more critical to better understanding customer browsing behavior across sites than accommodating the association in latent visit propensities.

Comparing the two model specifications under model 2, we find that the gamma mixing parameters are quite similar across the two models. Note, however, that the correlation parameter for the FGM hits its upper bound in the book category. It is worth emphasizing that this did not occur for any of the Sarmanov models we examined for either of our datasets. Moreover, model fit, as judged by LL or BIC, also strongly favors the Sarmanov specification. Therefore, we are left with a benchmark model that is dominated, both empirically and conceptually, by its Sarmanov counterpart.

Looking at the magnitudes of the correlation terms (and the improvements in the likelihood values as we move from model 1 toward model 4), it appears that the overall degree of dependence is moderately higher in music than in books. Perhaps this reflects the greater degree of competitive parity mentioned earlier. In both categories, however, these correlation terms (and likelihood gaps) are highly significant, offering strong support for the necessity of accommodating cross-site correlations.

5.2. Model Validation

Although we have shown that the proposed model performs well on a relative basis compared with various benchmark models, we have yet to show that

proposed model, but using the FGM bivariate exponential instead of the Sarmanov one. That is, we repeat all the steps discussed in Appendix B using FGM instead of Sarmanov.

Table 8 Model Results for Books

		Model 2				
		Model 1	FGM	Sarmanov	Model 3	Model 4
Amazon.com	r_A	0.4401	0.4743	0.4712	0.4372	0.4803
	α_A	0.2266	0.2369	0.2641	0.2161	0.2531
Barnesandnoble.com	r_B	0.2746	0.2795	0.2899	0.3118	0.3321
	α_B	0.3103	0.2919	0.2527	0.3023	0.2576
Correlation	ρ_τ	—	0.2500	0.1918	—	0.1944
	ρ_w	—	—	—	0.2066	0.2192
	LL	−4,260	−4,237	−4,221	−4,247	−4,208
	BIC	8,545	8,506	8,473	8,524	8,452

it performs sufficiently well on an absolute basis. An in-sample model validation of summary statistics helps demonstrate whether the proposed model properly captures the key behavioral aspects of browsing behavior across multiple sites.

Our objective here is to see how well the model results match the summary statistics discussed in §3. First, we examine the observed patterns of coincidence over time, discussed earlier and shown in Figure 3. Using the parameters from the proposed models, we simulate a new dataset and compare its coincidence patterns with the actual data. To provide some contrast, we perform this analysis for both the complete model and the independent one.

Figure 5 shows the comparisons. In general, the proposed model performs fairly well, especially over the longer timespans (and it dominates the independent model). Yet the model clearly underestimates the amount of coincident visits that occur within a short period of time. This gap is particularly acute for same-day coincidence.

A natural model extension to address this issue would be to introduce the concept of “hard-core comparison shoppers,” that is, customers who automatically go to one site the same day that they visit the other site. We could assume that there exists a fraction of p customers who visit websites in the manner of the proposed model, and the remaining $(1 - p)$ who visit both websites within the same day. If we were to allow for this “spike” in same-day shopping, we

conjecture that the curve for the complete model in Figure 5 would track the actual data far better than the one shown.

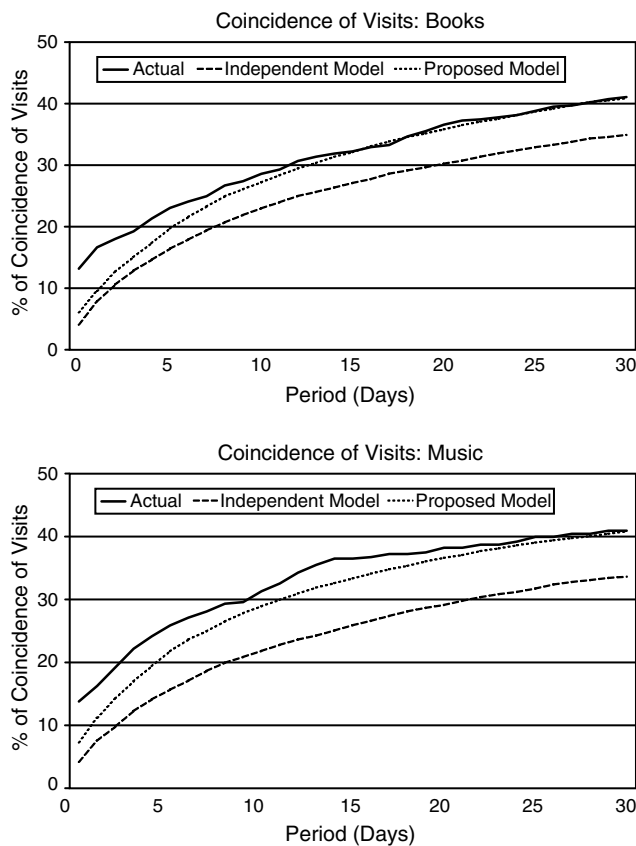
We do not pursue this extension here for several reasons: (1) This “spike” at zero begins to overcustomize the model in a way that may harm its generalizability and appropriateness for other domains (such as cross-channel visit patterns); (2) there are some questions about whether p should be specified as individual specific (i.e., applying to all visits for a given person) or visit specific among customers who visited sites multiple times; and (3) if we choose the latter option, we would need to consider the possibility of making p heterogeneous across the population, which would further complicate the model. In summary, this is a promising area for future research, but for now we simply accept the gaps seen in Figure 5 as a well-understood (and not especially critical) limitation of our model.

Finally, we investigate whether the simulated visit rates based on the proposed model reflect a similar degree of cross-site correlation as we see for the actual data. Using the same type of simulation process as described earlier in this section, we calculate the correlation between simulated visit rates in each product category. The null hypothesis is that the correlation between actual visit rates is equal to that between the simulated visit rates in a given product category. On the basis of t -test statistics, we cannot reject the null hypothesis in books and music at the 5% significance

Table 9 Model Results for Music

		Model 2				
		Model 1	FGM	Sarmanov	Model 3	Model 4
CDNOW.com	r_A	0.2097	0.2141	0.2305	0.2382	0.2887
	α_A	0.1658	0.1636	0.1899	0.1670	0.2086
Musicboulevard.com	r_A	0.2478	0.2477	0.2693	0.2824	0.2890
	α_B	0.2416	0.2287	0.2219	0.2407	0.2088
Correlation	ρ_τ	—	0.2130	0.2477	—	0.2435
	ρ_w	—	—	—	0.3041	0.2991
	LL	−3,067	−3,063	−3,034	−3,051	−3,021
	BIC	6,158	6,157	6,099	6,133	6,080

Figure 5 Model Validation: Coincidence of Visits



level. Thus, we are capturing the observed correlation pattern quite well. From a diagnostic standpoint, however, these correlations are not very meaningful; the parameters shown in Tables 8 and 9 are far better indicators of the true interrelationships between each pair of sites.

6. The Zero Class: Cross-Site Customer Acquisition

One of the principal motivations in building our cross-site browsing model is the ability to better understand—and potentially target—customers in the zero class. This group is the set of panelists who have been active in the product category at some previous point in the observation period but have not visited the site of interest during that time. As shown in Tables 2 and 3, the zero class accounts for the vast majority of panelists in our datasets: 76% for books and 83% for music. This is a natural focus of customer-acquisition efforts for online retailers, who have struggled with high costs (and meager outcomes) in performing this vital task. Unlike prior research that has dealt with the zero-class problem based solely on the data within a given source (Ehrenberg 1964, Morrison 1969, Morrison and Schmittlein 1981),

our research sheds new light on the issue by demonstrating how integrated databases can help improve understanding of new customer arrivals across sites.

We approach the zero-class problem in three ways. First is an aggregate analysis across a split sample: We predict the overall number of “period 1” zero-class members for a given site who will visit the site in “period 2.” Second, we derive and illustrate two different conditional expectation formulas to predict future visiting behavior at the individual level—one is for reach (i.e., the likelihood of visiting the previously nonvisited site at least once in a future period), and one is for frequency (i.e., the expected number of visits in the future period). Both of these conditional expectation formulas can help the marketing manager properly target valuable customers. Finally, we use these conditional expectations to validate the proposed model’s performance in predicting the acquisition of new customers.

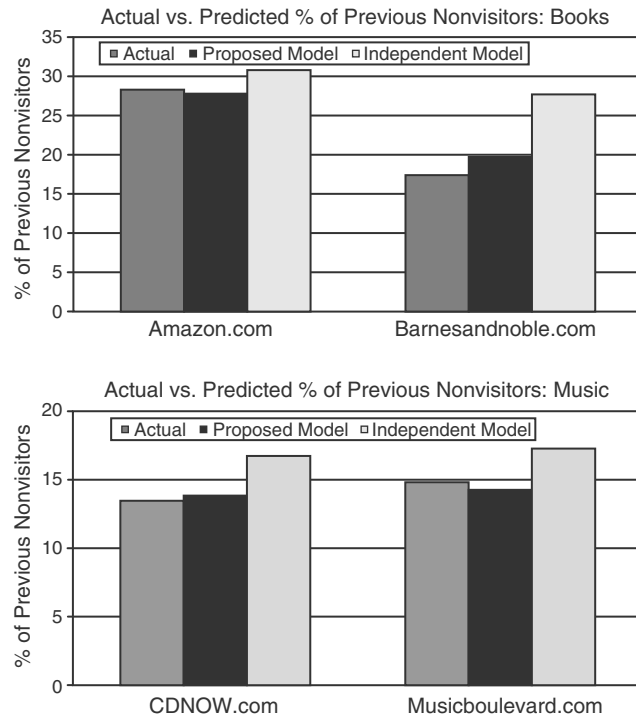
For all of the analyses conducted in this part of the paper, we use a longitudinal holdout sample. We re-estimate the relevant models (proposed and independent specifications) using only the first four months of data (period 1) for model calibration, and we save the second four months (period 2) for use as a holdout sample. We employ the simulation procedure described earlier to forecast the number of period 1 zero-class visitors at a specific site who visit the site in period 2.

6.1. Aggregate Analysis for the Zero Class

In this section, we examine the holdout forecast at the aggregate level. This is a relatively simple test and does not provide any specific information about which potential customers to target. But it is a useful way to further validate the proposed model and it highlights some of the biases and competitive elements that have been mentioned throughout the paper. See Figure 6 for a comparison of how well the proposed model and independent model perform in predicting the total number of period 2 buyers who were members of the zero class in period 1.

For all four sites, there is a significant improvement in forecasting by accounting for both associations in browsing patterns. The proposed model is clearly better at predicting the number of new buyers than the independent model. For each model, we test the null hypothesis—that its percentage of new buyers is equal to that of the actual data. For the proposed model, this null hypothesis cannot be rejected (at the 5% significance level) at each site in both product categories. For the independent model, however, the null hypothesis can be rejected at the 5% significance level at each site in both product categories, except for Amazon.com.

Beyond these formal tests, several interesting diagnostics emerge. In every case, the independent model

Figure 6 Actual Versus Predicted Percentage of the Zero Class in Period 1

overpredicts the number of previous nonvisitors. This result can be directly linked to the outcome of our earlier simulations, in which we observed that the failure to account for correlations in coincidence will lead to overestimates of the arrival rates. Thus, managers who rely on independent (site-centric) models may be painting an overly rosy picture of their abilities to naturally attract new customers.

In general, it is beneficial for all sites to use integrated visit-timing information, but the extent of these benefits is not the same across all sites. A further investigation of Figure 6 sheds light on which sites may benefit more by combining customer information across online retailers. For instance, in the books category, who benefits more from the combined database, Amazon.com or Barnesandnoble.com? This figure provides a clear answer: Barnesandnoble.com gains much more than Amazon.com (i.e., its forecast of customer acquisition is greatly improved) in moving from the independent model to the correlated one. It is logical that Amazon.com should realize less benefit compared to Barnesandnoble.com because the gigantic online bookseller already has enough information for customers who have browsed in the product category. In other words, Barnesandnoble.com learns a lot more from tapping into the Amazon.com customer histories than the other way around.

In contrast, the bottom panel of the figure shows a much greater degree of parity among the online music retailers. This result is quite plausible because, as shown in Table 1, these firms are fairly similar in terms of the size of their customer bases.

6.2. Conditional Expectations

We now approach the zero-class problem at the individual level by examining the following two questions: For a given individual who never visited a specific website (say, site A), but visited the other site (say, site B) in period 1, what is the expected likelihood for her to visit site A at least once in period 2? And, more specifically, what is her expected visit rate at site A in period 2? Without loss of generality, we focus on predicting a customer's future behavior at site A . We derive two analytical expressions of conditional expectation at site A in period 2, according to both the independent and the proposed model. These conditional expectations of future behavioral patterns can then be combined to make better inferences and predictions about the customer's visiting behavior across sites.

We will address these questions by considering whether this consumer was a recent or frequent visitor at site B in period 1. We examine whether, from site A 's perspective, recent visitors at site B are more valuable than frequent visitors. Therefore, this research provides a way to evaluate the "goodness" of consumers at the individual level by capturing these cross-site browsing patterns.

More formally, there are three elements—the observed calibration period ($T - t_0^i$) and the frequency of repeat visits at each site (A_{ji} and B_{ji})—that represent sufficient information for the independent model from the i th customer's visiting pattern. The proposed model requires two additional elements: the time of the last visit at each site (\underline{A}_j and \underline{B}_j in Appendix B) from the integrated database. We will calculate conditional expectations at site A in period 2 by varying the number of visits at site B (B_{ji}) and the time of last visit at site B (\underline{B}_j) in period 1.

We first show the conditional expectation formulas for the independent model, and in Appendix C we derive the equivalent expressions for the proposed model. Using the well-known expressions discussed in papers such as Morrison and Schmittlein (1981), the expected rate of visits at site A in period 2 is

$$\begin{aligned}
 E_A^I(\lambda_A^i | \text{Information}) &= \int_0^\infty \lambda_A^i \cdot g(\lambda_A^i | \text{Information}) d\lambda_A^i \\
 &= \frac{r_A + A_{ji}}{\alpha_A + T - t_0^i}, \quad (12)
 \end{aligned}$$

and the expected likelihood of no visit to site A in period 2 is

$$\begin{aligned} E_A^I(e^{-\lambda_A^i(T^*-T)} | \text{Information}) \\ &= \int_0^\infty e^{-\lambda_A^i(T^*-T)} \cdot g(\lambda_A^i | \text{Information}) d\lambda_A^i \\ &= \left(\frac{\alpha_A + T - t_0^i}{\alpha_A + T^* - t_0^i} \right)^{r_A + A_i}, \end{aligned} \quad (13)$$

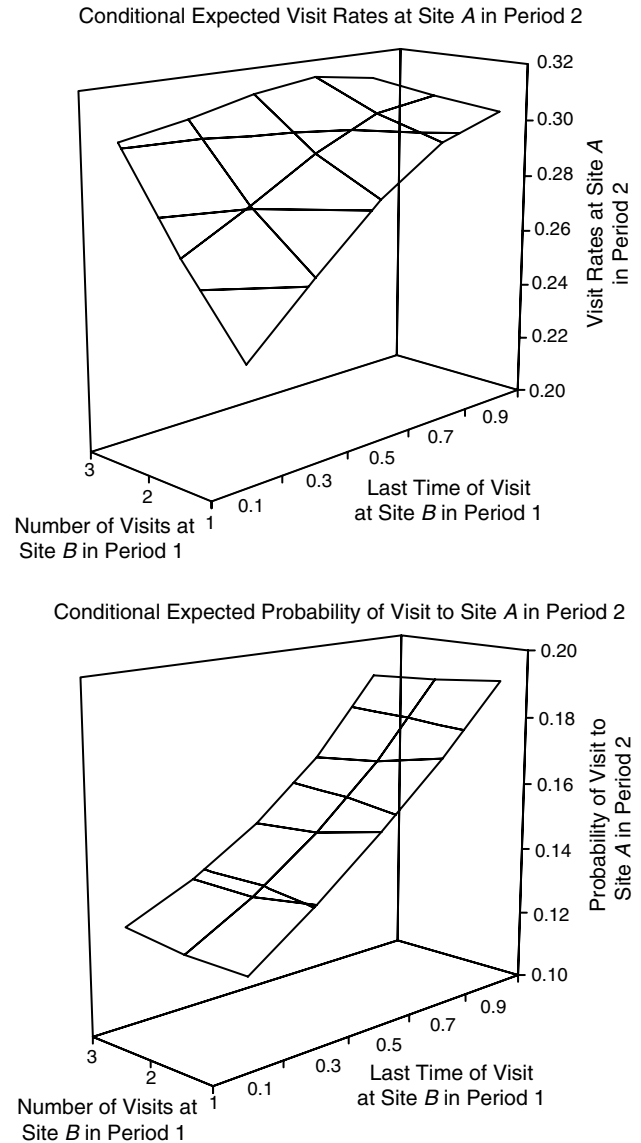
where λ_A^i is individual i 's latent visit rate at site A , (r_A, α_A) is a set of gamma parameters estimated in the independent model, T is the end of the observed calibration period, and T^* is the end of the future prediction period. To derive the expected probability of (at least) a visit to site A in period 2, we naturally subtract Equation (13) from the total probability, 1. Both conditional expressions of the independent model are well known. No conditional expectation of the independent model includes any information from site B in period 1, for instance, B_{i1} or \bar{B}_{i1} , and therefore is constant, regardless of the i th individual's visiting pattern at site B in period 1.

We discuss each conditional expectation of the proposed model in Appendix C. The method of deriving the conditional expectations for the proposed model is the same as for the independent model, except that the correlated bivariate distribution must be used. Each conditional expectation of the proposed model has the same basic form as in the independent model, but it is multiplied by a term that accounts for associations in the coincidence of visits and in latent visit rates. Unlike the constant conditional expectations of the independent model, the conditional expectations for the proposed model vary depending upon the information of the visit frequency and visit timing at site B in period 1. As noted earlier when discussing the likelihood function, each conditional expectation of the proposed model collapses down to that of the independent model if both types of cross-site dependence are set equal to zero.

In the following analysis, we use the parameter estimates for the book category estimated over the first four months. We rescale the calibration period from (t_0^i, T) to $(0, 1)$ for convenience. For the independent model, the expected rate of visits at site A (i.e., Amazon.com) in period 2 is a constant value (0.326) for every combination of the visit frequency and the time of the last visit at site B in period 1. As shown in the top panel of Figure 7, however, the conditional expectation for the proposed model varies depending upon the information on individual-level browsing behavior at site B in period 1.

The top panel of the figure shows some expected patterns—as well as some interesting interactions—between the recency and frequency of past visits at site B and the likely number of future visits to site A .

Figure 7 Conditional Expectations



The most valuable customers for site A to target, not surprisingly, are those who have visited site B very recently and frequently. Beyond this extreme point, we observe that for a frequent visitor to site B , recency is largely unimportant; the conditional expectations for the number of future visits to site A remain roughly constant with changes in recency. An equivalent story can be told for a very recent visitor: A customer who visited site B very recently has virtually the same expected future visit rate (at site A), regardless of the number of past visits he made to site B .

For past visitors who are not high on one of these dimensions, there are clear tradeoffs between recency and frequency that site A should take into account. For instance, among consumers who visited site B a long time ago, there exists a dramatic difference in

the expected rate of future (site *A*) visits by infrequent visitors than by frequent visitors. Customers with relatively high frequency but poor recency may have been heavy users of site *B* who defected for some reason.

In the lower panel of Figure 7, we see markedly different patterns for the probability of making at least one visit to site *A* based on past behavior at site *B*. Recency is far more important than frequency, which seems to be largely irrelevant in predicting whether a past site *B* customer will eventually visit site *A*. The recency effects are quite large: The probability of visiting site *A* is more than 50% higher for recent site *B* visitors compared to those who have not been there for a long time. This figure contrasts sharply with the constant visit probability of 0.216 that is associated with the independent model (reflecting, once again, the overestimate that arises due to model misspecification).

Taken together, both of these conditional expectations can be of great help for the marketing manager at site *A* in better targeting potential customers. They also hint at some tradeoffs that managers need to make between the desired reach versus frequency of their targeting efforts.

6.3. Empirical Validation of Conditional Expectations

In this section, we perform a final set of model validations to test the accuracy of the conditional expectations derived and discussed in the previous section. The goal is quite simple, but also very important: Do the period 2 conditional expectations for the members of the period 1 zero class at site *A* (Amazon.com) fall in line with their actual behavior? This is the strictest validation test of all, since it reflects individual-level variability in a holdout setting (as opposed to the aggregate, in-sample nature of tests, such as the one shown in Figure 5).

Table 10 shows the average number of actual period 2 visits for the members of the period 1 zero class and compares them to the estimated averages derived from the independent and proposed models. In each of the four cases, the estimated mean from the proposed model is closer to the actual figure than the corresponding estimate from the independent model.

Table 10 Conditional Expectations

	Actual visits	Independent model	Proposed model
Amazon.com	0.4717	0.4913	0.4626
Barnesandnoble.com	0.2609	0.3093	0.2802
CDNOW.com	0.3137	0.3451	0.2951
Musicboulevard.com	0.1875	0.2521	0.2039

Perhaps more noticeable (and interesting) is the bias associated with the independent model. As mentioned numerous times throughout the paper, we expect that the independent model will consistently inflate the estimated visit rates at each site, leading to the excessively high values shown in the table. Using the Wilcoxon signed rank test, we note that all four of the paired differences between the actual numbers and estimates from the independent model are statistically significant (in each case, the test yields $p < 0.001$). These results add to a mounting pile of evidence that this bias is genuine and quite meaningful.

In sharp contrast, the proposed model shows no evidence of any bias (two of the four estimates are above the actual mean, and two are below). More importantly, the sizes of these deviations are quite small. For three of the sites the Wilcoxon test yields $p > 0.2$. In the case of Amazon.com, however, the test suggests the existence of a significant difference ($p < 0.01$), but the actual size of this deviation is quite small (it is a fraction of the size of the biases arising from the independent model).

It is interesting to point out that the Amazon.com bias is actually an underestimate. In other words, previous nonvisitors are coming to the site in greater numbers than the model would predict. This observation could provide Amazon.com with useful diagnostics about dynamics that may be taking place at that site. Nevertheless, the fact that our static model can capture individual-level forecasts so well across the full set of sites is a very encouraging sign about its validity and managerial usefulness.

In the next section we identify several model extensions that might be able to address this issue, as well as other possible areas of improvement for the proposed model.

7. Conclusions and Future Research

This research focuses on combining Internet clickstream data from multiple online retailers. We provide a general framework for leveraging information (and explaining customer behavior) for browsing, which can also be applied to other settings, such as cross-channel and cross-business unit activities. To achieve this goal, we proposed two different sources of association in cross-site browsing patterns and examined them using a multivariate timing mixture model with closed-form analytic expressions. The Sarmanov family of multivariate distributions, new to the field of marketing, proved to be an invaluable asset, helping to overcome analytic hurdles faced by other researchers who have examined other types of multivariate shopping patterns (e.g., Chintagunta and Haldar 1998).

Using Internet clickstream data collected by Media Metrix, we demonstrated that this general (yet parsimonious) model offers significantly superior performance compared to the naïve independent model. A careful simulation showed that a failure to account for the two sources of association pattern not only leads to poor fits and forecasts, but also generates systematically biased parameter estimates. Chief among these simulation findings is the importance of allowing for coincidence in cross-site visit arrivals. A clear bias emerged in the form of overstated arrival rates if a positive arrival timing correlation is ignored (and vice versa for a negative correlation).

Even outside of the simulation study, the consequences of these biased estimates were evident throughout our empirical analysis, as we saw the independent model predicting excessively high visit rates.

Beyond the general improvement in model capabilities, this research sheds new light on the important topic of cross-site customer acquisition. We showed how the zero class (i.e., previous nonvisitors) for one site can be better understood by using an integrated database of past behavior across online retailers. In particular, we derived two analytical expressions for the conditional expectations at a given site in period 2 by considering whether a consumer was a recent or frequent visitor at the other site in period 1. These conditional expectations could be of great help for marketing managers in better targeting consumers with various marketing programs.

7.1. Data Availability

The concept of combining information across a set of online retailers is extremely timely. Industry experts are emphasizing the need for strategies built upon formal data integration (see, e.g., Forrester Research, Inc. 2002), and virtually all leading data vendors (including comScore Networks and Nielsen NetRatings) now offer services that allow clients to obtain data from multiple firms that compete within a particular industry. While most of these services primarily utilize aggregate data, these research vendors are willing to sell customized slices of disaggregate data to interested clients. As this market becomes increasingly competitive (and as clients become increasingly sophisticated and demanding), the “submarket” for session-level data is likely to become quite large.

We also see a number of industry-specific initiatives that feature the use of cross-site data. These include firms such as Autometrics for the auto industry and BigChampagne for digital music services. Perhaps the most visible example is an industry coalition named FAST (Future of Advertising Stakeholders) that is attempting to set industry standards for the measurement of online media audiences. A major focus of

this effort is to push firms toward user-centric measurement. One of the explicit objectives of FAST is to explain cross-site reach and frequency patterns, which is one of the main applications highlighted earlier.

Finally, in addition to all of these research-oriented initiatives, there are also a number of commercial ventures that actively use cross-site browsing behavior to target particular customers based on their sequence of visits across two or more sites. Probably the best known among these firms is the Gator Corporation, which has emerged into a successful enterprise with an impressive array of clients.

The bottom line here is that cross-site browsing data are available from a wide variety of sources, and there is clear commercial interest in expanding the scope of these activities as well as the analytic rigor associated with it.

7.2. Limitations and Future Research

Since this research is among the first attempts to investigate customer browsing behavior across online retailers using Internet clickstream data, we have kept the model as simple as possible to highlight the key phenomena that we have identified. Naturally, there are several limitations in the proposed model that should be acknowledged and perhaps addressed in future research. For instance, we implicitly assume that an individual is aware of both sites after his or her initial visit to either site in a given product category. Therefore, the observed period for the model estimation ($T - t_0^i$) is the same for both websites.

In addition, the proposed model does not consider the impact of marketing mix variables on visit-timing behavior within or across sites. This is a limitation of the Media Metrix data that is becoming less problematic with improvements in clickstream data-collection technology. As suitable explanatory variables become available, the model presented here can be easily extended to bring in the methods laid out by Gupta (1991), who carefully demonstrated the correct way of bringing time-varying covariates into a multiple-event timing model.

The addition of covariates can open up other types of multivariate benchmark models for comparison, such as vector ARMA models (Montgomery and Moe 2002). Not only is it worthwhile to compare the empirical performance of different classes of models, but it will be interesting to see if the same types of biases and managerial diagnostics continue to emerge.

As noted briefly earlier, the proposed model can also be extended to accommodate browsing behavior across more than two sites. The Sarmanov approach is by no means limited to the bivariate setting, and the necessary extensions to broader competitive contexts need not be overly burdensome. From a practical standpoint, the jump from two to three (or more) sites

is far less daunting than the initial step from the univariate case to the bivariate setting. We discuss some of these issues in a technical appendix available from the authors and encourage other researchers to follow up on them.

Another area for future research is allowing for nonstationarity in the multivariate timing mixture model. Moe and Fader (2004a) demonstrate the prevalence of nonstationarity in site-visit behavior and develop an extended exponential-gamma model to deal with it. Some of this seemingly nonstationary behavior may arise due to cross-site effects that are ignored in a site-centric model. (Our model's generally strong performance in holdout settings is far better than that of the static benchmark models employed by Moe and Fader.) To the extent that we detect any nonstationarity, it might be interesting to see if "evolving" behavior occurs evenly across sites, or if there is a compensatory pattern (i.e., speeding up visits at one site while slowing down visits at another).

Finally, it will be useful to consider applications beyond Internet browsing behavior alone. The model can be used for online or offline purchase behavior, as well as some of the other CRM issues mentioned at the outset of the paper, such as the use of multiple channels or "touchpoints" by the customers of a particular firm. For these behavioral issues and others, we hope that researchers will begin to adopt broader views of customer shopping patterns to keep our model-building activities on a par with—or potentially ahead of—the data-collection capabilities that make such models possible.

Acknowledgments

The authors would like to thank comScore Media Metrix for access to their data, Eric J. Johnson for assembling and sharing this particular dataset, and the reviewers for an unusually constructive set of comments and suggestions. This paper stems from the first author's dissertation work, and he extends special thanks to his dissertation committee (Eric Bradlow, Steve Hoch, Abba Krieger, and Jagmohan Raju) for their helpful feedback and encouragement.

Appendix A. Hazard Functions

When working with the proposed multivariate timing model, it is important to note that the hazard functions are no longer constant because the hazard rate at a given site is a function of intervisit times at its competing site as well as at its own site. It is therefore necessary to account for the elapsed time since the last visit at each site in constructing the correlated timing process across sites.

For expositional clarity in this appendix alone, we define time as the intervisit time between successive visits to each site, as opposed to the actual calendar time that we use elsewhere in the paper. Suppose that these intervisit times follow an (independent univariate) exponential distribution at each site. We denote by t_s the random variable

associated with the intervisit times at site s and $f_s(t_s) = \lambda_s e^{-\lambda_s t_s}$ for $s = A, B$ the univariate exponential densities. Let $f_{AB}^I(t_A, t_B)$ be the bivariate exponential distribution with the independence assumption, which is the product of two univariate exponential densities; that is, $f_{AB}^I(t_A, t_B) = f_A(t_A) \times f_B(t_B)$. The hazard function for site A , $h_A^I(t_A, t_B)$, is given by

$$h_A^I(t_A, t_B) = \frac{\int_{t_B}^{\infty} f_{AB}^I(t_A, x_B) dx_B}{\int_{t_A}^{\infty} \int_{t_B}^{\infty} f_{AB}^I(x_A, x_B) dx_B dx_A} = \lambda_A. \quad (14)$$

Likewise, the hazard function for site B , $h_B^I(t_A, t_B)$, is λ_B . This represents the typical memoryless property of the exponential timing process.

Now consider the hazard functions of the Sarmanov bivariate exponential distribution, $f_{AB}^D(t_A, t_B)$, which accounts for the dependence in visit timing across sites. The hazard function for site A , $h_A^D(t_A, t_B)$, is given by:

$$\begin{aligned} h_A^D(t_A, t_B) &= \frac{\int_{t_B}^{\infty} f_{AB}^D(t_A, x_B) dx_B}{\int_{t_A}^{\infty} \int_{t_B}^{\infty} f_{AB}^D(x_A, x_B) dx_B dx_A} \\ &= \lambda_A \times \left[\left(1 + \rho_{\tau} \frac{(1 + \lambda_A)^2}{\lambda_A} (1 + \lambda_B) \left(e^{-t_A} - \frac{\lambda_A}{1 + \lambda_A} \right) \cdot (e^{-t_B} - 1) \right) / (1 + \rho_{\tau} (1 + \lambda_A) (1 + \lambda_B)) \cdot (e^{-t_A} - 1) (e^{-t_B} - 1) \right]. \quad (15) \end{aligned}$$

The hazard function for site B , $h_B^D(t_A, t_B)$, can be derived in a similar manner. The hazard function at each site is no longer constant in the presence of correlation. When this correlation is ignored (or equal to zero), these hazard functions collapse into those of the separate univariate densities.

The hazard function derived above addresses the following question: "Given the fact that no arrivals have occurred at either website over a period of time t_A and t_B since the last arrival at each site, what is the probability that an arrival will be realized at site A in the next instant?" (Of course we recognize that the hazard function cannot be interpreted as a probability, but the main point here should be clear.)

But in some managerial settings, there may be interest in the *conditional* hazard function, that is, "Given that the last arrival occurred at site B , what is the probability that an arrival will be realized at site A in the next instant?" This hazard function⁸ is given by

$$\begin{aligned} h_{A|B}^D(t_A | t_B) &= \frac{f_{A|B}^D(t_A | t_B)}{\int_{t_A}^{\infty} f_{A|B}^D(x_A | t_B) dx_A} \\ &= \lambda_A \times \left[\left(1 + \rho_{\tau} \frac{(1 + \lambda_A)^2}{\lambda_A} \frac{(1 + \lambda_B)^2}{\lambda_B} \left(e^{-t_A} - \frac{\lambda_A}{1 + \lambda_A} \right) \cdot \left(e^{-t_B} - \frac{\lambda_B}{1 + \lambda_B} \right) \right) / \left(1 + \rho_{\tau} (1 + \lambda_A) \cdot \frac{(1 + \lambda_B)^2}{\lambda_B} (e^{-t_A} - 1) \left(e^{-t_B} - \frac{\lambda_B}{1 + \lambda_B} \right) \right) \right]. \quad (16) \end{aligned}$$

⁸ We would like to thank an anonymous reviewer for pointing out (and deriving) this conditional hazard function.

Once again, in the absence of any correlation in the cross-site timing behavior, this expression collapses into λ_A by itself.

Appendix B. Likelihood Function of the Complete Model

We construct the overall likelihood function of the complete model for parameter estimation purposes. To do so, we first define the notation, which illustrates the data required to estimate the models. We next discuss two main components in the proposed multivariate timing mixture model: timing process and heterogeneity distribution. Finally, we derive the unconditional distribution to build the overall likelihood function of the complete model.

B.1. Notation

For individual i , let t_j^i , $j = 1, 2, \dots, J_i$, be the random time of repeat visits at either site, where t_0^i is the day of her initial visit at either site and T is the end of the observed calibration period. J_i is the total number of repeat visits at either site individual i makes. Note that it is necessary to take a superposition on the visit times combined across sites in order to draw the information of browsing patterns in constructing the multivariate timing model.⁹

Let I_j , $j = 1, 2, \dots, J_i$, be the random indicator, where $I_j = 1$, if the j th repeat visit occurs at site A for individual i and 0 otherwise. Finally, define $\underline{A}_j = \{k: I_k = 1, I_{k+1} = I_{k+2} = \dots = I_j = 0, \text{ for } k \leq j\}$. That is, \underline{A}_j is the random indicator of the last visit at site A before (or on) the j th repeat visit made by individual i . In a similar way, $\underline{B}_j = \{k: I_k = 0, I_{k+1} = I_{k+2} = \dots = I_j = 1, \text{ for } k \leq j\}$. These random indicators of past visit patterns denote elapsed time since the last visit at each site.

The independent model only requires the information of t_j^i and I_j from the combined browsing data, which is equivalent to the information of the observed calibration period (i.e., $T - t_0^i$) and the total frequency of repeat visits at each site (i.e., A_{J_i} and B_{J_i}). Thus, the independent model does not require the visit timing information. To fit the correlated bivariate timing process, we need to extract \underline{A}_j and \underline{B}_j as well as t_j^i and I_j from the integrated database. Thus, the proposed approach uses more detailed information across sites.

B.2. Timing Process

We categorize visits by the current visit and the last visit pattern at each site. Three kinds of current visits are: visit to site A , visit to site B , and no visit at either site because of right censoring. To accommodate the notion of nonconstant hazard rates in the bivariate exponential timing process, it

is necessary to consider past visit history using \underline{A}_j and \underline{B}_j . Hence, there are nine different types of visit:

Previous visit	The j th repeat visit		
	Visit site A	Visit site B	No visit
No visit	$V_{\cdot A}^i$	$V_{\cdot B}^i$	$S_{\cdot N}^i$
Visit site A	V_{AA}^i	V_{AB}^i	S_{AN}^i
Visit site B	V_{BA}^i	V_{BB}^i	S_{BN}^i

The first row in the table describes different types of the first repeat visit at site A ($V_{\cdot A}^i$), site B ($V_{\cdot B}^i$), and no realized visit at either site ($S_{\cdot N}^i$) made by individual i . The second and last rows construct the timing process conditional on the previous visit history. The second row assumes that the i th individual made the $(j-1)$ th repeat visit at site A . It implies that $t_{j-1}^i - t_{B_j}^i$ is an elapsed time where no arrival has been realized at site B . In a similar vein, the $(j-1)$ th repeat visit at site B implies that $t_{j-1}^i - t_{A_j}^i$ is an elapsed time where no arrival has been realized at site A . Each cell can be read as it stands. For instance, the cell S_{AN}^i can be read as follows: Given the $(j-1)$ th repeat visit at site A , that is, conditional on $t_{j-1}^i - t_{B_j}^i$, no arrival has been realized because of right censoring. The conditional joint timing densities of each cell (the first main component in the timing mixture model) are discussed next.

Let us derive the (conditional) joint timing density of the first repeat visit at site A and site B and of no visit at either site for individual i , that is, $V_{\cdot A}^i$, $V_{\cdot B}^i$, and $S_{\cdot N}^i$, respectively. Let $f_s(t_j^i; \lambda_s^i) = \lambda_s^i e^{-\lambda_s^i(t_j^i - t_{j-1}^i)}$, $S_s(t_j^i - t_{j-1}^i; \lambda_s^i) = e^{-\lambda_s^i(t_j^i - t_{j-1}^i)}$, $\phi_s(t_j^i - t_{j-1}^i; \lambda_s^i) = e^{-(t_j^i - t_{j-1}^i)} - \lambda_s^i / (1 + \lambda_s^i)$, and $\bar{\phi}(t_j^i - t_{j-1}^i) = e^{-(t_j^i - t_{j-1}^i)} - 1$ for $s = A, B$. If the first repeat visit made by individual i occurs at site A , the conditional distribution of $V_{\cdot A}^i$ is given by

$$\begin{aligned} f(t_{1(I_1=1)}^i; \lambda_A^i, \lambda_B^i, \rho_\tau) \\ = \int_{t_1^i}^{\infty} f(t_1^i, x_2^i; \lambda_A^i, \lambda_B^i, \rho_\tau) dx_2^i = f_A(t_1^i; \lambda_A^i) \times S_B(t_1^i - t_0^i; \lambda_B^i) \\ \cdot \left\{ 1 + \rho_\tau \frac{(1 + \lambda_A^i)^2}{\lambda_A^i} (1 + \lambda_B^i) \phi_A(t_1^i - t_0^i; \lambda_A^i) \bar{\phi}(t_1^i - t_0^i) \right\}. \quad (17) \end{aligned}$$

This joint density can be understood in a simple and intuitive way. Observing the first repeat visit at site A , which means that no arrival has occurred at site B by t_1^i , leads to an exponential density for site A ($f_A(t_1^i; \lambda_A^i)$) and a survivor function for site B ($S_B(t_1^i - t_0^i; \lambda_B^i)$) with a combined expression of dependence in the coincidence of visits. In a similar way, if the first repeat visit is realized at site B for individual i , the conditional distribution of $V_{\cdot B}^i$ is given by

$$\begin{aligned} f(t_{1(I_1=0)}^i; \lambda_A^i, \lambda_B^i, \rho_\tau) \\ = \int_{t_1^i}^{\infty} f(x_1^i, t_1^i; \lambda_A^i, \lambda_B^i, \rho_\tau) dx_1^i = S_A(t_1^i - t_0^i; \lambda_A^i) \times f_B(t_1^i; \lambda_B^i) \\ \cdot \left\{ 1 + \rho_\tau (1 + \lambda_A^i) \frac{(1 + \lambda_B^i)^2}{\lambda_B^i} \bar{\phi}(t_1^i - t_0^i) \phi_B(t_1^i - t_0^i; \lambda_B^i) \right\}. \quad (18) \end{aligned}$$

⁹ Note that in order to make the model development as clear as possible, this notation (and the concept of superposition as a whole) was not used earlier in the paper. These technical issues are required for model estimation, but do not affect the underlying logic of the model's derivation.

This expression may be interpreted as the opposite of Equation (17), that is, a survivor function for site A ($S_A(t_1^i - t_0^i; \lambda_A^i)$) and an exponential distribution for site B ($f_B(t_1^i; \lambda_B^i)$) with a combined expression to capture dependence in visit timing between the two sites. Finally, when neither site is visited by individual i during the entire observed period after the initial visit at either site, the conditional distribution of S_N^i can be written as follows:

$$\begin{aligned} & S(T - t_0^i; \lambda_A^i, \lambda_B^i, \rho_\tau) \\ &= \int_T^\infty \int_T^\infty f(x_1^i, x_2^i; \lambda_A^i, \lambda_B^i, \rho_\tau) dx_1^i dx_2^i \\ &= S_A(T - t_0^i; \lambda_A^i) \times S_B(T - t_0^i; \lambda_B^i) \\ &\quad \cdot \{1 + \rho_\tau(1 + \lambda_A^i)(1 + \lambda_B^i)\bar{\phi}(T - t_0^i)^2\}. \end{aligned} \quad (19)$$

Now no exponential density can be observed because no arrival has occurred at either site, that is, a survivor function for each site with a combined correlated expression.

From the second repeat visit made by individual i , it is necessary to condition the information of the past visit patterns to consider nonconstant hazard rates in the bivariate exponential timing process. Derivations of the bivariate timing process of the second row in the table, that is, V_{AA}^i , V_{AB}^i , and S_{AN}^i , are very similar to the joint timing density of the first row, except that it is required to condition the $(j-1)$ th repeat visit realized at site A in deriving $f(t_{j(l_j=1)}^i | t_{j-1}^i - t_{B_j}^i; \lambda_A^i, \lambda_B^i, \rho_\tau)$, $f(t_{j(l_j=0)}^i | t_{j-1}^i - t_{B_j}^i; \lambda_A^i, \lambda_B^i, \rho_\tau)$, and $S(T - t_{j(l_j=1)}^i | t_{j-1}^i - t_{B_j}^i; \lambda_A^i, \lambda_B^i, \rho_\tau)$. Finally, in constructing the bivariate timing process of the last row in the table, that is, V_{BA}^i , V_{BB}^i , and S_{BN}^i , it is required to condition the $(j-1)$ th repeat visit at site B in deriving $f(t_{j(l_j=1)}^i | t_{j-1}^i - t_{A_j}^i; \lambda_A^i, \lambda_B^i, \rho_\tau)$, $f(t_{j(l_j=0)}^i | t_{j-1}^i - t_{A_j}^i; \lambda_A^i, \lambda_B^i, \rho_\tau)$, and $S(T - t_{j(l_j=1)}^i | t_{j-1}^i - t_{A_j}^i; \lambda_A^i, \lambda_B^i, \rho_\tau)$.

B.3. Heterogeneity Distribution

We previously discussed the standard approach to obtain the likelihood function for a given individual. The first step is to specify the individual-level likelihood function for each site, conditional on that person's latent visit rate at that site. We then multiply all s_{h_i} exponential terms, times an additional term to account for the right censoring. Once this conditional likelihood function is constructed, we then integrate across all possible values of λ_s^i using the gamma mixing distribution to get the unconditional distribution for each site. Finally, the multiplication of unconditional likelihood for each site yields the simple bivariate model, which can be multiplied across N individuals to get the overall likelihood for parameter estimation purposes.

An alternative path that leads to the same result is to perform the gamma integration separately for each of the J_i bivariate exponential terms, times an additional term to account for the right censoring, and then multiply them together at the end. This involves the use of Bayes theorem to update our guess about each individual's value of λ_A^i and λ_B^i as each arrival occurs. Specifically, if someone's first repeat visit occurs at site A at time t_1^i , then

$$\begin{aligned} & g(\lambda_{A_2}^i, \lambda_{B_2}^i | \text{visit site } A \text{ at } t_1^i) \\ &= \text{gamma}(r_A + 1, \alpha_A + t_1^i - t_0^i, r_B, \alpha_B + t_1^i - t_0^i, \omega). \end{aligned}$$

While scale parameters in the gamma distribution (i.e., α_s for $s = A, B$) are always updated (i.e., $\alpha_s + t_1^i - t_0^i$) at each site, a shape parameter (i.e., r_s for $s = A, B$) in the gamma density is only updated as long as a visit is realized at a specific site (i.e., $r_s + 1$). If a visit has not occurred at a given site, the shape parameter remains the same. The bivariate gamma distribution governing latent visit rates for subsequent arrivals is as follows:

$$\begin{aligned} & g(\lambda_{A_{j+1}}^i, \lambda_{B_{j+1}}^i | \text{arrival at } t_j^i) \\ &= \text{gamma}\left(r_A + \sum_{k=1}^j I_k, \alpha_A + t_j^i - t_0^i, r_B + j \right. \\ &\quad \left. - \sum_{k=1}^j I_k, \alpha_B + t_j^i - t_0^i, \omega\right) \\ &\stackrel{\text{let}}{=} \text{gamma}(r_A^O, \alpha_A^O, r_B^O, \alpha_B^O, \omega). \end{aligned} \quad (20)$$

This updating procedure is used to get the unconditional distribution by integrating the conditional distribution of the bivariate timing process across all possible values of the latent visit rates.

B.4. Unconditional Distribution

We next derive the unconditional distribution of the bivariate timing process. To get the unconditional distribution of each type of visit, we integrate each conditional distribution of the bivariate timing process across all possible values of λ_A^i and λ_B^i using the bivariate gamma mixing distribution, which updates each individual's value of λ_A^i and λ_B^i as each arrival is realized. If the first repeat visit occurs at site A , the unconditional distribution of V_{AA}^i is given by

$$\begin{aligned} f(t_{1(l_1=1)}^i) &= \int_0^\infty \int_0^\infty (17) \times g(\lambda_A^i, \lambda_B^i; r_A, \alpha_A, r_B, \alpha_B, \omega) d\lambda_A^i d\lambda_B^i \\ &= \frac{r_A}{\alpha_A} \left(\frac{\alpha_A}{\alpha_A^N}\right)^{r_A^N} \left(\frac{\alpha_B}{\alpha_B^N}\right)^{r_B} \\ &\quad \cdot \left[\begin{aligned} & 1 + \rho_\tau h_{AB}^1(t_1^i; r_A^N, \alpha_A^N, r_B, \alpha_B^N) \\ & + \omega h_{AB}^2(t_1^i; r_A^N, \alpha_A^N, r_B, \alpha_B^N) \\ & + \rho_\tau \omega h_{AB}^3(t_1^i; r_A^N, \alpha_A^N, r_B, \alpha_B^N) \end{aligned} \right], \end{aligned} \quad (21)$$

where

$$\begin{aligned} h_{AB}^1 &= \left\{ e^{-(t_1^i - t_0^i)} \left(2 + \frac{\alpha_A^N}{r_A} + \frac{r_A^N}{\alpha_A^N} \right) - \left(1 + \frac{r_A^N}{\alpha_A^N} \right) \right\} \times \left\{ 1 + \frac{r_B}{\alpha_B^N} \right\} \{ e^{-(t_1^i - t_0^i)} - 1 \}, \\ h_{AB}^2 &= \left\{ \left(\frac{\alpha_A^N}{1 + \alpha_A^N} \right)^{r_A^N} - \left(\frac{\alpha_A}{1 + \alpha_A} \right)^{r_A} \right\} \times \left\{ \left(\frac{\alpha_B^N}{1 + \alpha_B^N} \right)^{r_B} - \left(\frac{\alpha_B}{1 + \alpha_B} \right)^{r_B} \right\}, \\ h_{AB}^3 &= \{ e^{-(t_1^i - t_0^i)} - 1 \} \left\{ \left(\frac{\alpha_A^N}{1 + \alpha_A^N} \right)^{r_A^N} \left[e^{-(t_1^i - t_0^i)} \left(2 + \frac{1 + \alpha_A^N}{r_A} + \frac{r_A^N}{1 + \alpha_A^N} \right) \right. \right. \\ &\quad \left. \left. - \left(1 + \frac{r_A^N}{1 + \alpha_A^N} \right) \right] - \left(\frac{\alpha_A}{1 + \alpha_A} \right)^{r_A} \right. \\ &\quad \left. \cdot \left[e^{-(t_1^i - t_0^i)} \left(2 + \frac{\alpha_A^N}{r_A} + \frac{r_A^N}{\alpha_A^N} \right) - \left(1 + \frac{r_A^N}{\alpha_A^N} \right) \right] \right\} \\ &\quad \cdot \left\{ \left[\left(\frac{\alpha_B^N}{1 + \alpha_B^N} \right)^{r_B} \left(1 + \frac{r_B}{1 + \alpha_B^N} \right) - \left(\frac{\alpha_B}{1 + \alpha_B} \right)^{r_B} \left(1 + \frac{r_B}{\alpha_B^N} \right) \right] \right\}, \end{aligned}$$

and $r_s^N = r_s + 1$ and $\alpha_s^N = \alpha_s + t_1^i - t_0^i$ for $s = A, B$. If the first repeat visit is realized at site B , the unconditional distribu-

tion of V_{AB}^i is given by

$$\begin{aligned} f(t_{1(l_1=0)}^i) &= \int_0^\infty \int_0^\infty (18) \times g(\lambda_A^i, \lambda_B^i; r_A, \alpha_A, r_B, \alpha_B, \omega) d\lambda_A^i d\lambda_B^i \\ &= \left(\frac{\alpha_A}{\alpha_A^N} \right)^{r_A} \frac{r_B}{\alpha_B} \left(\frac{\alpha_B}{\alpha_B^N} \right)^{r_B} \\ &\quad \cdot \left[1 + \rho_\tau h_{AB}^1(t_1^i; r_A, \alpha_A^N, r_B^N, \alpha_B^N) \right. \\ &\quad \left. + \omega h_{AB}^2(t_1^i; r_A, \alpha_A^N, r_B^N, \alpha_B^N) \right. \\ &\quad \left. + \rho_\tau \omega h_{AB}^3(t_1^i; r_A, \alpha_A^N, r_B^N, \alpha_B^N) \right], \end{aligned} \quad (22)$$

where

$$\begin{aligned} h_{AB}^1 &= \left\{ 1 + \frac{r_A}{\alpha_A^N} \right\} \{ e^{-(t_1^i - t_0^i)} - 1 \} \\ &\quad \cdot \left\{ e^{-(t_1^i - t_0^i)} \left(2 + \frac{\alpha_B^N}{r_B} + \frac{r_B^N}{\alpha_B^N} \right) - \left(1 + \frac{r_B^N}{\alpha_B^N} \right) \right\}, \\ h_{AB}^2 &= \left\{ \left(\frac{\alpha_A^N}{1 + \alpha_A^N} \right)^{r_A} - \left(\frac{\alpha_A}{1 + \alpha_A} \right)^{r_A} \right\} \\ &\quad \cdot \left\{ \left(\frac{\alpha_B^N}{1 + \alpha_B^N} \right)^{r_B} - \left(\frac{\alpha_B}{1 + \alpha_B} \right)^{r_B} \right\}, \\ h_{AB}^3 &= \left\{ \left[\left(\frac{\alpha_A^N}{1 + \alpha_A^N} \right)^{r_A} \left(1 + \frac{r_A}{1 + \alpha_A^N} \right) - \left(\frac{\alpha_A}{1 + \alpha_A} \right)^{r_A} \left(1 + \frac{r_A}{\alpha_A^N} \right) \right] \right. \\ &\quad \cdot \{ e^{-(t_1^i - t_0^i)} - 1 \} \left\{ \left(\frac{\alpha_B^N}{1 + \alpha_B^N} \right)^{r_B} \left[e^{-(t_1^i - t_0^i)} \left(2 + \frac{1 + \alpha_B^N}{r_B} + \frac{r_B^N}{1 + \alpha_B^N} \right) \right. \right. \\ &\quad \left. \left. - \left(1 + \frac{r_B^N}{1 + \alpha_B^N} \right) \right] - \left(\frac{\alpha_B}{1 + \alpha_B} \right)^{r_B} \right. \right. \\ &\quad \left. \left. \cdot \left[e^{-(t_1^i - t_0^i)} \left(2 + \frac{\alpha_B^N}{r_B} + \frac{r_B^N}{\alpha_B^N} \right) - \left(1 + \frac{r_B^N}{\alpha_B^N} \right) \right] \right\} \right\}, \end{aligned}$$

and $r_s^N = r_s + 1$ and $\alpha_s^N = \alpha_s + t_1^i - t_0^i$ for $s = A, B$. If no arrival occurs during the entire observed calibration period, the unconditional distribution of S_{AB}^i is given by

$$\begin{aligned} S(T - t_0^i) &= \int_0^\infty \int_0^\infty (19) \times g(\lambda_A^i, \lambda_B^i; r_A, \alpha_A, r_B, \alpha_B, \omega) d\lambda_A^i d\lambda_B^i \\ &= \left(\frac{\alpha_A}{\alpha_A^N} \right)^{r_A} \left(\frac{\alpha_B}{\alpha_B^N} \right)^{r_B} \\ &\quad \cdot \left[1 + \rho_\tau h_{AB}^1(t_1^i; r_A, \alpha_A^N, r_B, \alpha_B^N) \right. \\ &\quad \left. + \omega h_{AB}^2(t_1^i; r_A, \alpha_A^N, r_B, \alpha_B^N) \right. \\ &\quad \left. + \rho_\tau \omega h_{AB}^3(t_1^i; r_A, \alpha_A^N, r_B, \alpha_B^N) \right], \end{aligned} \quad (23)$$

where

$$\begin{aligned} h_{AB}^1 &= \left\{ \left(1 + \frac{r_A}{\alpha_A^N} \right) (e^{-(T - t_0^i)} - 1) \right\} \times \left\{ \left(1 + \frac{r_B}{\alpha_B^N} \right) (e^{-(T - t_0^i)} - 1) \right\}, \\ h_{AB}^2 &= \left\{ \left(\frac{\alpha_A^N}{1 + \alpha_A^N} \right)^{r_A} - \left(\frac{\alpha_A}{1 + \alpha_A} \right)^{r_A} \right\} \\ &\quad \cdot \left\{ \left(\frac{\alpha_B^N}{1 + \alpha_B^N} \right)^{r_B} - \left(\frac{\alpha_B}{1 + \alpha_B} \right)^{r_B} \right\}, \\ h_{AB}^3 &= (e^{-(T - t_0^i)} - 1) \left\{ \left(\frac{\alpha_A^N}{1 + \alpha_A^N} \right)^{r_A} \left(1 + \frac{r_A}{1 + \alpha_A^N} \right) \right. \\ &\quad \left. - \left(\frac{\alpha_A}{1 + \alpha_A} \right)^{r_A} \left(1 + \frac{r_A}{\alpha_A^N} \right) \right\} \end{aligned}$$

$$\cdot (e^{-(T - t_0^i)} - 1) \left\{ \left(\frac{\alpha_B^N}{1 + \alpha_B^N} \right)^{r_B} \left(1 + \frac{r_B}{1 + \alpha_B^N} \right) \right. \\ \left. - \left(\frac{\alpha_B}{1 + \alpha_B} \right)^{r_B} \left(1 + \frac{r_B}{\alpha_B^N} \right) \right\},$$

and $\alpha_s^N = \alpha_s + T - t_0^i$ for $s = A, B$. Similarly, derivations of the other unconditional distributions, $f(t_{j(l_j=1)}^i | t_{j-1}^i - t_{B_j}^i)$, $f(t_{j(l_j=0)}^i | t_{j-1}^i - t_{B_j}^i)$, $S(T - t_{l_i}^i | t_{l_i}^i - t_{B_{l_i}}^i)$, $f(t_{j(l_j=1)}^i | t_{j-1}^i - t_{A_j}^i)$, $f(t_{j(l_j=0)}^i | t_{j-1}^i - t_{A_j}^i)$, and $S(T - t_{l_i}^i | t_{l_i}^i - t_{A_{l_i}}^i)$ are straightforward.

B.5. Likelihood Function of the Proposed Model

On the basis of nine different unconditional distributions derived in the previous section, we can construct the overall likelihood function of the multivariate timing mixture model proposed in this research:

$$L_{AB}^D = \prod_{i=1}^N \prod_{j=1}^{l_i} \left[\begin{aligned} &[f(t_{1(l_1=1)}^i; r_A, \alpha_A, r_B, \alpha_B, \rho_\tau, \rho_\omega)]^{(V_{AA}^i)^1} \\ &[f(t_{1(l_1=0)}^i; r_A, \alpha_A, r_B, \alpha_B, \rho_\tau, \rho_\omega)]^{(V_{AB}^i)^1} \\ &[S(T - t_0^i; r_A, \alpha_A, r_B, \alpha_B, \rho_\tau, \rho_\omega)]^{(S_{AB}^i)^1} \\ &[f(t_{j(l_j=1)}^i | t_{j-1}^i - t_{B_j}^i; r_A, \alpha_A, r_B, \alpha_B, \rho_\tau, \rho_\omega)]^{(V_{AA}^i)^j} \\ &[f(t_{j(l_j=0)}^i | t_{j-1}^i - t_{B_j}^i; r_A, \alpha_A, r_B, \alpha_B, \rho_\tau, \rho_\omega)]^{(V_{AB}^i)^j} \\ &[S(T - t_{l_i}^i | t_{l_i-1}^i - t_{B_{l_i}}^i; r_A, \alpha_A, r_B, \alpha_B, \rho_\tau, \rho_\omega)]^{(S_{AB}^i)^j} \\ &[f(t_{j(l_j=1)}^i | t_{j-1}^i - t_{A_j}^i; r_A, \alpha_A, r_B, \alpha_B, \rho_\tau, \rho_\omega)]^{(V_{BA}^i)^j} \\ &[f(t_{j(l_j=0)}^i | t_{j-1}^i - t_{A_j}^i; r_A, \alpha_A, r_B, \alpha_B, \rho_\tau, \rho_\omega)]^{(V_{BB}^i)^j} \\ &[S(T - t_{l_i}^i | t_{l_i-1}^i - t_{A_{l_i}}^i; r_A, \alpha_A, r_B, \alpha_B, \rho_\tau, \rho_\omega)]^{(S_{BA}^i)^j} \end{aligned} \right], \quad (24)$$

where the superscripted indicator variable, for example, $(V_{AA}^i)^j$, is 1 if the j th repeat visit matches to the corresponding type of visit for individual i and 0 otherwise. The likelihood of the complete model collapses down to that of the simple bivariate model, once two types of cross-site dependencies are completely ignored.

Appendix C. Conditional Expectations

Without loss of generality, we derive two expressions for the conditional expectation at site A in period 2 in the proposed model: the expected rate of visits at site A in period 2 and the expected likelihood of at least a visit to site A in period 2. Since we consider customers who never visited site A but visited site B in period 1, note that the last visit was realized at site B in period 1. While three elements $(T - t_0^i, A_{l_i}$, and $B_{l_i})$ represent sufficient information from the i th customer's visiting pattern for the independent model, the proposed model requires two additional elements of the time of the last visit at each site (A_j and B_j) from the combined database.

C.1. Expected Rate of Visits at Site A in Period 2

Given the information of the i th customer browsing pattern at sites A and B in period 1, in the proposed model, the

expected rate of visits at site A in period 2 is

$$\begin{aligned} E_A^D(\lambda_A^i | \text{Information}) \\ = \int_0^\infty \lambda_A^i \cdot g(\lambda_A^i | \text{Information}) d\lambda_A^i \\ = \left(\frac{r_A + A_{j_i}}{\alpha_A + T - t_0^i} \right) \left(\frac{1 + \rho_\tau h_{AB}^{N_1} + \omega h_{AB}^{N_2} + \rho_\tau \omega h_{AB}^{N_3}}{1 + \rho_\tau h_{AB}^{D_1} + \omega h_{AB}^{D_2} + \rho_\tau \omega h_{AB}^{D_3}} \right), \end{aligned} \quad (25)$$

where

$$\begin{aligned} h_{AB}^{D_1} &= \left\{ \left(1 + \frac{r_A^O}{\alpha_A^N} \right) (e^{-(T-t_{j_i}^i)} - 1) \right\} \times \left\{ \left(1 + \frac{r_B^O}{\alpha_B^N} \right) (e^{-(T-t_{\Delta j_i}^i)} - 1) \right\}, \\ h_{AB}^{D_2} &= \left\{ \left(\frac{\alpha_A^N}{1 + \alpha_A^N} \right)^{r_A^O} - \left(\frac{\alpha_A^O}{1 + \alpha_A^O} \right)^{r_A^O} \right\} \\ &\quad \cdot \left\{ \left(\frac{\alpha_B^N}{1 + \alpha_B^N} \right)^{r_B^O} - \left(\frac{\alpha_B^O}{1 + \alpha_B^O} \right)^{r_B^O} \right\}, \\ h_{AB}^{D_3} &= (e^{-(T-t_{j_i}^i)} - 1) \left\{ \left(\frac{\alpha_A^N}{1 + \alpha_A^N} \right)^{r_A^O} \left(1 + \frac{r_A^O}{1 + \alpha_A^N} \right) \right. \\ &\quad \left. - \left(\frac{\alpha_A^O}{1 + \alpha_A^O} \right)^{r_A^O} \left(1 + \frac{r_A^O}{\alpha_A^N} \right) \right\} \\ &\quad \cdot (e^{-(T-t_{\Delta j_i}^i)} - 1) \left\{ \left(\frac{\alpha_B^N}{1 + \alpha_B^N} \right)^{r_B^O} \left(1 + \frac{r_B^O}{1 + \alpha_B^N} \right) \right. \\ &\quad \left. - \left(\frac{\alpha_B^O}{1 + \alpha_B^O} \right)^{r_B^O} \left(1 + \frac{r_B^O}{\alpha_B^N} \right) \right\}, \\ h_{AB}^{N_1} &= \left\{ \left(1 + \frac{r_A^N}{\alpha_A^N} \right) (e^{-(T-t_{j_i}^i)} - 1) \right\} \\ &\quad \cdot \left\{ \left(1 + \frac{r_B^O}{\alpha_B^N} \right) (e^{-(T-t_{\Delta j_i}^i)} - 1) \right\}, \\ h_{AB}^{N_2} &= \left\{ \left(\frac{\alpha_A^N}{1 + \alpha_A^N} \right)^{r_A^N} - \left(\frac{\alpha_A^O}{1 + \alpha_A^O} \right)^{r_A^O} \right\} \\ &\quad \cdot \left\{ \left(\frac{\alpha_B^N}{1 + \alpha_B^N} \right)^{r_B^O} - \left(\frac{\alpha_B^O}{1 + \alpha_B^O} \right)^{r_B^O} \right\}, \\ h_{AB}^{N_3} &= (e^{-(T-t_{j_i}^i)} - 1) \left\{ \left(\frac{\alpha_A^N}{1 + \alpha_A^N} \right)^{r_A^N} \left(1 + \frac{r_A^N}{1 + \alpha_A^N} \right) \right. \\ &\quad \left. - \left(\frac{\alpha_A^O}{1 + \alpha_A^O} \right)^{r_A^O} \left(1 + \frac{r_A^N}{\alpha_A^N} \right) \right\} \\ &\quad \cdot (e^{-(T-t_{\Delta j_i}^i)} - 1) \left\{ \left(\frac{\alpha_B^N}{1 + \alpha_B^N} \right)^{r_B^O} \left(1 + \frac{r_B^O}{1 + \alpha_B^N} \right) \right. \\ &\quad \left. - \left(\frac{\alpha_B^O}{1 + \alpha_B^O} \right)^{r_B^O} \left(1 + \frac{r_B^O}{\alpha_B^N} \right) \right\}, \end{aligned}$$

and $r_s^N = r_s^O + 1$ and $\alpha_s^N = \alpha_s^O + T - t_{j_i}^i$ for $s = A, B$.

C.2. Expected Likelihood of at Least One Visit to Site A in Period 2

Given the information of the i th customer browsing pattern at sites A and B in period 1, in the proposed model, the

expected likelihood of no visit to site A in period 2 is

$$\begin{aligned} E_A^D(e^{-\lambda_A^i(T^*-T)} | \text{Information}) \\ = \int_0^\infty e^{-\lambda_A^i(T^*-T)} \cdot g(\lambda_A^i | \text{Information}) d\lambda_A^i \\ = \left(\frac{\alpha_A + T - t_0^i}{\alpha_A + T^* - t_0^i} \right)^{r_A + A_{j_i}} \\ \cdot \left(\frac{1 + \rho_\tau h_{AB}^{N_1} + \omega h_{AB}^{N_2} + \rho_\tau \omega h_{AB}^{N_3}}{1 + \rho_\tau h_{AB}^{D_1} + \omega h_{AB}^{D_2} + \rho_\tau \omega h_{AB}^{D_3}} \right), \end{aligned} \quad (26)$$

where

$$\begin{aligned} h_{AB}^{D_1} &= \left\{ \left(1 + \frac{r_A^O}{\alpha_A^N} \right) (e^{-(T-t_{j_i}^i)} - 1) \right\} \\ &\quad \cdot \left\{ \left(1 + \frac{r_B^O}{\alpha_B^N} \right) (e^{-(T-t_{\Delta j_i}^i)} - 1) \right\}, \\ h_{AB}^{D_2} &= \left\{ \left(\frac{\alpha_A^N}{1 + \alpha_A^N} \right)^{r_A^O} - \left(\frac{\alpha_A^O}{1 + \alpha_A^O} \right)^{r_A^O} \right\} \\ &\quad \cdot \left\{ \left(\frac{\alpha_B^N}{1 + \alpha_B^N} \right)^{r_B^O} - \left(\frac{\alpha_B^O}{1 + \alpha_B^O} \right)^{r_B^O} \right\}, \\ h_{AB}^{D_3} &= (e^{-(T-t_{j_i}^i)} - 1) \left\{ \left(\frac{\alpha_A^N}{1 + \alpha_A^N} \right)^{r_A^O} \left(1 + \frac{r_A^O}{1 + \alpha_A^N} \right) \right. \\ &\quad \left. - \left(\frac{\alpha_A^O}{1 + \alpha_A^O} \right)^{r_A^O} \left(1 + \frac{r_A^O}{\alpha_A^N} \right) \right\} \\ &\quad \cdot (e^{-(T-t_{\Delta j_i}^i)} - 1) \left\{ \left(\frac{\alpha_B^N}{1 + \alpha_B^N} \right)^{r_B^O} \left(1 + \frac{r_B^O}{1 + \alpha_B^N} \right) \right. \\ &\quad \left. - \left(\frac{\alpha_B^O}{1 + \alpha_B^O} \right)^{r_B^O} \left(1 + \frac{r_B^O}{\alpha_B^N} \right) \right\}, \\ h_{AB}^{N_1} &= \left\{ \left(1 + \frac{r_A^O}{\alpha_A^{N*}} \right) (e^{-(T-t_{j_i}^i)} - 1) \right\} \\ &\quad \cdot \left\{ \left(1 + \frac{r_B^O}{\alpha_B^N} \right) (e^{-(T^*-t_{\Delta j_i}^i)} - 1) \right\}, \\ h_{AB}^{N_2} &= \left\{ \left(\frac{\alpha_A^{N*}}{1 + \alpha_A^{N*}} \right)^{r_A^O} - \left(\frac{\alpha_A^O}{1 + \alpha_A^O} \right)^{r_A^O} \right\} \\ &\quad \cdot \left\{ \left(\frac{\alpha_B^N}{1 + \alpha_B^N} \right)^{r_B^O} - \left(\frac{\alpha_B^O}{1 + \alpha_B^O} \right)^{r_B^O} \right\}, \\ h_{AB}^{N_3} &= (e^{-(T-t_{j_i}^i)} - 1) \left\{ \left(\frac{\alpha_A^{N*}}{1 + \alpha_A^{N*}} \right)^{r_A^O} \left(1 + \frac{r_A^O}{1 + \alpha_A^{N*}} \right) \right. \\ &\quad \left. - \left(\frac{\alpha_A^O}{1 + \alpha_A^O} \right)^{r_A^O} \left(1 + \frac{r_A^O}{\alpha_A^{N*}} \right) \right\} \\ &\quad \cdot (e^{-(T^*-t_{\Delta j_i}^i)} - 1) \left\{ \left(\frac{\alpha_B^N}{1 + \alpha_B^N} \right)^{r_B^O} \left(1 + \frac{r_B^O}{1 + \alpha_B^N} \right) \right. \\ &\quad \left. - \left(\frac{\alpha_B^O}{1 + \alpha_B^O} \right)^{r_B^O} \left(1 + \frac{r_B^O}{\alpha_B^N} \right) \right\}, \end{aligned}$$

$\alpha_s^N = \alpha_s^O + T - t_{j_i}^i$ and $\alpha_s^{N*} = \alpha_s^O + T^* - t_{j_i}^i$ for $s = A, B$, and T^* is the end of the future prediction period.

References

- Bucklin, Randolph E., Catarina Sismeiro. 2003. A model of web site browsing behavior estimated on clickstream data. *J. Marketing Res.* **40**(August) 249–267.
- Chintagunta, Pradeep K., Sudeep Haldar. 1998. Investigating purchase timing behavior in two related product categories. *J. Marketing Res.* **35**(February) 43–53.
- Ehrenberg, Andrew S. C. 1964. Estimating the proportion of loyal buyers. *J. Marketing Res.* **1**(February) 56–59.
- Farlie, D. J. G. 1960. The performance of some correlation coefficients for a general bivariate distribution. *Biometrika* **47**(December) 307–323.
- Forrester Research, Inc. 2002. *Getting the Retail Technology Advantage*. Forrester Research, Inc., Cambridge, MA.
- Gupta, Sunil. 1991. Stochastic models of interpurchase time with time-dependent covariates. *J. Marketing Res.* **28**(February) 1–15.
- Iyer, Ganesh, Amit Pazgal. 2003. Internet shopping agents: Virtual co-location and competition. *Marketing Sci.* **22**(1) 85–106.
- Johnson, Eric J., Steven Bellman, Gerald L. Lohse. 2003. Cognitive lock-in and the power law of practice. *J. Marketing* **67**(2) 62–75.
- Johnson, Eric J., Wendy W. Moe, Peter S. Fader, Steven Bellman, Gerald L. Lohse. 2004. On the depth and dynamics of online search behavior. *Management Sci.* **50**(3) 326–335.
- Johnson, Norman L., Samuel Kotz. 1975. On some generalized Farlie-Gumbel-Morgenstern distributions. *Comm. Statist.* **4**(5) 415–427.
- Johnson, Norman L., Samuel Kotz. 1977. On some generalized Farlie-Gumbel-Morgenstern distributions—II: Regression, correlation, and further generalizations. *Comm. Statist. Theory Methods* **A6**(6) 485–496.
- Kotz, Samuel, N. Balakrishnan, Norman L. Johnson. 2000. *Continuous Multivariate Distributions*, 2nd ed. Volume 1: Models and Applications. John Wiley & Sons, New York.
- Lee, Mei-Ling Ting. 1996. Properties and applications of the Sarmanov family of bivariate distributions. *Comm. Statist. Theory Methods* **25**(6) 1207–1222.
- Moe, Wendy W., Peter S. Fader. 2004a. Capturing evolving visit behavior in clickstream data. *J. Interactive Marketing* **18**(1) 5–19.
- Moe, Wendy W., Peter S. Fader. 2004b. Dynamic conversion behavior at e-commerce sites. *Management Sci.* **50**(3) 326–335.
- Montgomery, Alan L., Wendy W. Moe. 2002. Should record companies pay for radio airplay? Investigating the relationship between music sales and radio airplay using vector autoregressive-moving average models. Working paper, Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, PA.
- Morrison, Donald G. 1969. Conditional trend analysis: A model that allows for nonusers. *J. Marketing Res.* **6**(August) 342–346.
- Morrison, Donald G., David C. Schmittlein. 1981. Predicting future random events based on past performance. *Management Sci.* **27**(September) 1006–1023.
- Padmanabhan, Balaji, Zhiqiang Zheng, Steven O. Kimbrough. 2001. Incomplete data in personalization: What you don't know can hurt. *Proc. of the ACM Internat. Conf. on Knowledge Discovery and Data Mining*, San Francisco, CA, 154–164.
- Saha, Atanu, Lynette Hilton. 1997. Expo-power: A flexible hazard function for duration data models. *Econom. Lett.* **54** 227–233.
- Sarmanov, O. V. 1966. Generalized normal correlation and two-dimensional frechet classes. *Soviet Math.—Doklady*, **7** 596–599 [English translation; Russian original in *Doklady Akademii Nauk SSSR*, 108 (1966)].
- Seetharaman, P. B., Pradeep K. Chintagunta. 2003. The proportional hazard model for purchase timing: A comparison of alternative specifications. *J. Bus. Econom. Statist.* **21**(3) 1–15.
- Shop.org. 2003. *The State of Online Retailing* 6.0.
- Swift, Ronald S. 2001. *Accelerating Customer Relationships*. Prentice Hall, Upper Saddle River, NJ.