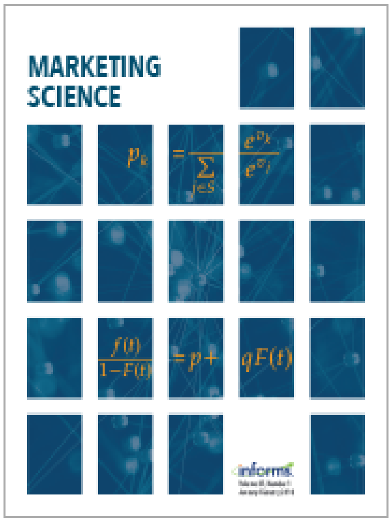


Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Investigating Purchase Conversion by Uncovering Online Visit Patterns

Chang Hee Park, Young-Hoon Park

MARKETING SCIENCE
Vol. 35, No. 6, November–December 2016, pp. 894–914
ISSN 0732-2399 (print) | ISSN 1526-548X (online)

https://doi.org/10.1287/mksc.2016.0990

Investigating Purchase Conversion by Uncovering Online Visit Patterns

Chang Hee Park
School of Management, Binghamton University, State University of New York, Binghamton, New York 13902, cpark@binghamton.edu

Young-Hoon Park
Samuel Curtis Johnson Graduate School of Management, Cornell University, Ithaca, New York 14853, yp34@cornell.edu

This research aims to understand and predict online customers' store visit and purchase behaviors. To this end, we develop a model that accounts for different patterns of online store visits at the individual level. Given the latency of visit patterns, we employ a changepoint modeling framework and statistically infer them using a Bayesian approach. The inferences obtained are then used to examine the effects of visit patterns on purchase dynamics across store visits. Using Internet clickstream data at an online retailer, we find that online store visit patterns tend to be clustered with significant variation across customers in terms of the number and size of visit clusters as well as the visit frequencies, both within and between clusters. Furthermore, the conversion rates vary significantly, depending on store visit patterns, such that they tend to be higher at later visits within a visit cluster, compared with earlier visits. The proposed model thereby offers superior fit and predictive performance than benchmark models that ignore clustered visit patterns and their impact on purchase behavior. We demonstrate the model's ability to better identify prospective customers by utilizing their visit patterns, which can assist marketers in scoring customers and making targeting decisions across individuals for marketing activity.

Data, as supplemental material, are available at https://doi.org/10.1287/mksc.2016.0990.

Keywords: online shopping behavior; timing models; changepoint models; pattern analysis; Bayesian estimation
History: Received: May 30, 2014; accepted: January 22, 2016; Fred Feinberg served as the senior editor and Scott Neslin served as associate editor for this article. Published online in *Articles in Advance* July 14, 2016.

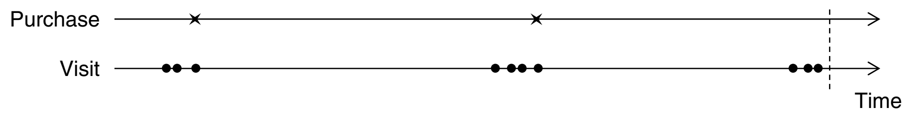
1. Introduction

Since the Internet emerged as a major medium for commercial purposes, online shopping has continued to see phenomenal growth. In the United States, with a 10% compound annual growth rate, online retail sales are expected to reach $480 billion by 2019, up from $334 billion in 2015, such that they will account for 11% of overall retail sales (Forrester Research 2015). Despite the continuing success of e-commerce channels, most online retailers convert only a small percentage of their customer visits to purchases. Naturally, understanding and predicting visit-to-purchase conversion behavior have been of great importance to marketers at e-commerce sites, because small changes in the conversion rate can lead to considerable increases in revenues.

A recent survey reports that more than 50% of sales at online stores are driven by repeat customers (Forrester Research 2014). Recognizing this, an increasing number of firms seek to provide customers with personalized content and services based on their past shopping behavior, a promising way of increasing purchase conversions (e.g., Hauser et al. 2009). A significant aid to such efforts is a more accurate prediction of customers' visit timing and purchase likelihood. Marketing researchers have developed sophisticated models to examine and predict conversion behavior (e.g., Montgomery et al. 2004, Sismeiro and Bucklin 2004, Hauser et al. 2009). Yet, most of them focus on purchase behavior *within* a given visit, conditional on a customer's arrival at the retailer's website. In this research, we investigate online shopping dynamics *across* store visits and identify customers' visit patterns as an important source of predictive information about their visit and purchase behaviors.

Consider a series of visit events to an online store, made by an individual customer over a time period. The visit history can be characterized as a point process on a one-dimensional (time) space. According to the literature on pattern recognitions and spatial analysis, everything that has a location in a space inevitably creates or contributes to patterns (e.g., Upton and Fingelton 1985, Porter and Howell 2012). These patterns, in point processes, can be broadly classified as uniform, random, or clustered (e.g., Cliff and Ord 1981, Miller and Han 2009). A pattern of points exhibits uniformity when the series of events

Figure 1 An Example of Online Shopping Behavior



occurs with the same intervals on the space. Random patterns have no underlying regularity in the locations of points, such that the intervals between events are determined in a stochastic manner and thus the position of a point is not affected by the positions of others. A clustered pattern refers to a process that exhibits local concentrations of events in close spatial or temporal proximity, with each cluster separated by empty or less dense patterns of events. Of these three types, clustered patterns attract our attention because of their prevalence in online customer behavior (e.g., Zhang et al. 2015).

To explore clustered patterns and their importance in examining online customer behavior, we consider an example of customer shopping patterns in Figure 1. Figure 1 illustrates the sequence of visits (Internet sessions) and purchase conversions by a customer at the online retailer considered in this research. A • symbol indicates that the customer visited the online store, and a × symbol indicates that she made a purchase. The customer made 10 visits and two purchases during the data period. The first 3 visits, which she made in the early part of the data period, have relatively short intervisit times. Because these visits are in close temporal proximity and the concentrations are separated by an elapsed period of no visits, the 3 visits could be considered a cluster of visit events. Similarly, the next 4 visits by the customer could constitute another visit cluster. After all, her visit process exhibits clustered patterns with three clusters of visit events during the time period. We also find various clustered patterns of store visits by other customers in our database.

Using extant models in the marketing literature that study the time series of visit or purchase data (e.g., Gupta 1991, Jain and Vilcassim 1991, Fader et al. 2005), marketing managers at online stores can predict when their customers will visit next and whether they will make purchases during those future visits. A key question arises here as to whether (and how) these predictions could be improved by taking customers' store visit patterns into account. For the customer shown in Figure 1, a clear pattern emerges. Within each visit cluster described above, her first visit to the online store is immediately followed by a few more visits with relatively short intervisit times, and she tends to make a purchase at the last visit in the cluster. After the observation period ends (vertical dashed line), therefore, the customer may continue making clustered visits and likely purchase toward the end of the visit cluster. If so, the subsequent intervisit time would be much smaller if her next visit belongs to the last visit cluster, compared to the case in which she leaves the visit cluster and her next visit initiates a new visit cluster. Likewise, the purchase likelihood may still systematically vary, depending on the patterns of future visits. Clearly, the predictions on visit-to-purchase conversion behavior would differ from those drawn without the consideration of the visit patterns.

The objective of this research is to better understand and predict online customers' store visit and purchase behaviors. To this end, we develop a model that flexibly captures different patterns of online store visits and accounts for their effects on purchase dynamics. Generalizing extant multievent timing models built on the stochastic process of customer events (e.g., Ehrenberg 1988, Gupta 1988), our visit model accounts for the notion that a customer's visit process may consist of multiple visit clusters, each with different visit rates within and between clusters and a varying number of visit events that constitute a visit cluster. Because the start and the end of visit clusters are unobserved, because of their latency in the visit process, we employ and extend a dynamic changepoint modeling framework (e.g., Fader et al. 2004) and statistically infer individual customers' visit patterns using a Bayesian approach. The inferred visit patterns are then used in examining and predicting their purchase behavior as well as store visit behavior. Our model thereby allows us to capture various time-varying patterns of online store visits and their effects on purchase dynamics. We also account for the effects of marketing activity on visit and purchase behaviors, the possibility of latent customer attrition (e.g., Schmittlein et al. 1987), and different sources of customer heterogeneity in online shopping behavior.

We apply our model to customers' visit and purchase data from an online fashion retailer. Our results suggest that online store visit patterns tend to be clustered such that customers' visit timing process consists of several visit clusters with considerably higher visit rates within clusters and lower visit rates between clusters. Using the inferred formation of visit clusters, we find that the conversion rates vary substantially, depending on the size of a visit cluster and the location of a visit event in a cluster. Overall, the conversion rates are higher at later visits within a cluster, compared with earlier visits. In line with this, we find that taking clustered visit patterns into account plays an important role in predicting customers' purchase decision at a given visit. Our model thereby offers superior fit and performance in predicting online purchase as well as visit behavior, compared with benchmark models that fail to consider the clustered visit patterns that underlie customers' shopping behavior at the online retailer.

Our model and Bayesian estimation approach also provide a set of useful summaries on customers' store visit patterns. Specifically, it allows us to obtain (1) the number of visits per cluster, (2) the intervisit times within a cluster, (3) the time length of a cluster, (4) the intervisit times between clusters, and (5) the number of visit clusters in a given time period, at the individual customer level. We find significant variation across customers in terms of the number and size of visit clusters as well as the visit frequencies, both within and between visit clusters. These cluster-based inferences at the individual level can assist an online retailer in tailoring marketing communications across customers. For example, for high-clustering customers who tend to make several visits within clusters (versus low-clustering customers), when their first few clustered visits end with no purchase, the online retailer could benefit by communicating with the customers, with customized offers and services based on their browsing activities (e.g., products searched or viewed) at those prior visits, as they are likely to come back to the retailer shortly and make additional visits within clusters. We also demonstrate that, compared to existing approaches, our model enables us to better identify prospective visitors and buyers, which helps managers improve customers scoring and targeting decisions across individuals for marketing activity. Given technical difficulties and large costs for a complete real-time customization of website content and services (e.g., Telang et al. 2004), the improved predictions of visit timing and purchase likelihood afforded by our model could assist managers' customization and resource allocation efforts.

The remainder of this article is organized as follows. In Section 2, we provide an overview of prior literature related to our work and discuss the contribution of the research to the literature. Section 3 discusses the conceptual background that motivates our modeling approach. Section 4 presents a formal specification of our model. In Section 5, we provide a description of the data used in our empirical analysis and discuss model results. We conclude with directions for future research in Section 6.

2. Related Literature

This study is related to prior research on online customer behavior and multievent timing models. We briefly review relevant literature on both streams of research and discuss the contribution of our work relative to them.

Understanding online customer behavior has been of great interest to marketing researchers. Prior studies investigate browsing behavior in websites (e.g., Bucklin and Sismeiro 2003, Park and Fader 2004, Telang et al. 2004, Danaher 2007). Several researchers link browsing behavior to purchase behavior by looking at the sequence of pages viewed (e.g., Montgomery et al. 2004) or tasks completed (e.g., Sismeiro and Bucklin 2004) within a site visit. Others address the cumulative effects of website visits between purchase conversions (e.g., Moe and Fader 2004), the effects of online banner advertising (e.g., Manchanda et al. 2006), and the cognitive styles of shoppers (e.g., Hauser et al. 2009). While these studies provide useful insights about conversion behavior, they typically model only purchase behavior within a given visit, by assuming store visits to be exogenous data rather than stochastic events. The applicability of the models is accordingly restricted to purchase events conditional on visits. By comparison, we jointly model customers' visit timing process and purchase decisions upon visits in an integrated framework and examine how a series of visits leads to purchase conversions.

From a methodological perspective, this study draws on prior research that has investigated customers' visit and/or purchase timing behavior. Much research in this domain has relied on the assumption of stationary exponential timing processes (e.g., Schmittlein et al. 1985, Ehrenberg 1988). Several researchers have extended exponential timing models to consider the effects of time-varying explanatory variables (e.g., Gupta 1988, 1991), unobserved nonstationarity in the customer arrival processes (e.g., Fader et al. 2004, Schweidel and Fader 2009), latent customer attrition (e.g., Schmittlein et al. 1987, Fader et al. 2005), and alternative parametric assumptions on baseline timing processes (e.g., Jain and Vilcassim 1991, Seetharaman and Chintagunta 2003). We contribute to this body of literature by developing a multievent timing model that allows for the clustered visit patterns of online customers and shopping dynamics within and between clusters. In particular, generalizing extant models in customer base analysis (e.g., Schmittlein et al. 1987, Fader et al. 2005, Abe 2009), our model allows customers' visit rate to vary dynamically depending on the formation of latent visit clusters in the visit process as well as observed time-varying covariates, while accounting for latent customer attrition. Our model also contributes to the hidden Markov model (HMM) framework (e.g., Montgomery et al. 2004, Netzer et al. 2008, Schweidel et al. 2009) through a flexible high-order dependence of behavioral outcomes on latent states. Finally, our model extends existing changepoint models (e.g., Fader et al. 2004, Schweidel and Fader 2009) by allowing the probability of a changepoint occurring to vary dynamically at the individual level.

Nascent literature on the patterns of customer behavior examines some notable aspects of data characteristics and their role in predicting future behavior. Schwartz et al. (2014) study the interplay between the summary statistics of longitudinal data and the performance of alternative models. Their resulting decision tree assists the choice of model frameworks for the analysis of customer data. Zhang et al. (2015) is closely related to our work in that they look into the clumpiness (i.e., clustering phenomenon) of data patterns and relate it to customer behavior. Their research discusses the deficiency in RFM (recency, frequency, and monetary values) measures in summarizing customer transaction history, and shows that the clumpiness is a useful predictor of customer lifetime values. Our work substantially differs from Zhang et al. (2015) in terms of the ways of utilizing clustered data patterns as well as the research focus. Zhang et al. (2015) compute the clumpiness measures, fully conditional on data, and then correlate the measures with the outcome variables of interest. By contrast, our research models the data generation process itself and thus enables us to predict the data patterns. From a substantive viewpoint, to our knowledge, our research is the first to examine whether predictions of online shopping behavior could be improved by taking store visit patterns into consideration.

3. Conceptual Background

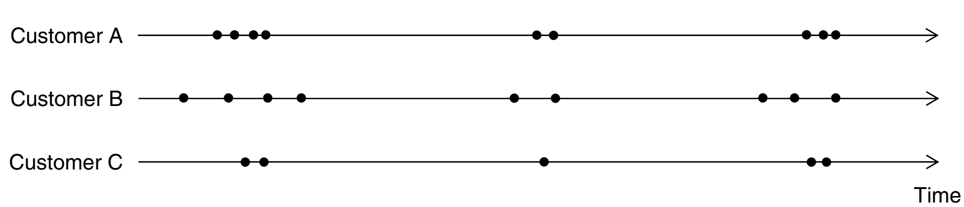
Central to our objective is to develop a model that flexibly captures various patterns of online store visits and their impacts on purchase behavior. In particular, our focus is on the consideration of clustered visit patterns, because they are common online and, more important, could be informative to better understand customer behavior at e-commerce sites. Clustered data patterns have been observed and studied in various empirical contexts such as epidemiology, criminology, seismology, and finance as well as sports (e.g., Gilovich et al. 1985; Murray 2000; Bernasco and Nieuwbeerta 2005; Zhang et al. 2013, 2015). In the literature, the clustering phenomenon is also referred to as data clumpiness or the "hot hand" effect to reflect the observation that success breeds success and failure breeds failure.

Clustered data patterns exhibit bursts of events in close vicinity to one another, with each cluster separated by empty or less dense patterns of events. The notion is accordingly relative rather than absolute because patterns can be more or less clustered in comparison to one another (e.g., Zhang et al. 2013). We take on the challenge to characterize and infer clustered visit patterns of online customers with two unique sources in the visit timing process: one for the intervisit times within and between clusters and the other for the number of visit events that constitute a visit cluster. We illustrate these concepts with an example.

Figure 2 shows why it is important to consider the two different sources to describe clustered visit patterns. Consider the visit behaviors of customer A and customer B. They appear to be identical in terms of the number of visits to the online store during the data period. In addition, on the basis of longer intervisit times between clusters of visit events, both customers formed three visit clusters. However, once we consider the intervisit times within clusters, we see different visit behaviors between the customers. For customer A, her first visit to the online store is quickly followed by a few more visits, with short intervisit times. By comparison, customer B does not seem to have a high level of visit frequency within clusters, leading to relatively long intervisit times within clusters. They also differ in terms of the intervisit times between clusters: customer B's intervisit times between clusters are shorter than those of customer A. Given the intervisit times within and between clusters, the visit patterns of customer A are more clustered than those of customer B.

We next consider customer A and customer C to describe the second source of clustered visit patterns. As described, customer A makes at least a few visit events within clusters. By comparison, customer C makes only one or two visits per cluster. Hence, the visit patterns of customer C are less clustered than those of customer A. This characterization of clustered patterns is based on the number of visit events that constitute a visit cluster. Therefore, to properly model the online customers' visit process, it is imperative

Figure 2 **Examples of Clustered Visit Patterns**



that we account for both sources of clustered visit patterns.

Before we formally lay out our model, one relevant question to ask is: What underlies clustered visits online? Several reasons may exist, but the clustered patterns of store visits can be largely attributed to low transportation and browsing costs required to visit an online store, a unique characteristic of online shopping (e.g., Bakos 1997, Johnson et al. 2004). Behavioral research on the consumer decision-making process suggests that a purchase decision is typically preceded by stages of information search and evaluation, during which individuals undertake various prepurchase activities, such as comparing alternative products and reading product reviews (e.g., Urbany et al. 1989, Moorthy et al. 1997). One of the main factors that influence the amount of these preparatory activities is the cost of performing the tasks. As it is essentially costless for a customer to visit and browse a retailer's website, such low transportation costs online allow and encourage shoppers to make several visits to the online store before making a purchase decision (e.g., Moe 2003, Moe and Fader 2004). In addition, a common feature offered at many e-commerce sites that allows customers to save items in their online shopping basket for future visits may also contribute to the multivisit tendency for a single purchase decision online. This series of frequent visits are then represented as a cluster of visit events in the customer's visit timing process.

Yet, online customers' multivisit tendency for a single purchase decision may not be the only explanation for the clustered patterns of their store visits. Other reasons may include seasonality in demand, cyclic changes in customer preferences, budget and time constraints of customers, and retailers' marketing activity, to name but a few. Along these lines, we recognize that the underlying reasons for clustered visits, which are beyond the scope of this research, are often related to customers' intrinsic characteristics that are unobservable and also different across individual customers. This leads us to take a stochastic approach to model the store visit process, while accounting for the effects of available covariates that may affect the visit patterns.

4. Model Development

We propose an integrated modeling framework that allows for the inference of visit patterns to an online store and investigates their impacts on purchase behavior in a noncontractual setting. We begin by modeling customers' purchase decisions conditional on their visits to the online retailer. In Section 4.2, we build a model of visit timing behavior that characterizes store visit patterns. Section 4.3 discusses our accommodation of unobserved heterogeneity across customers and the correlation structure of model components. We then formulate the likelihood function of the proposed model in Section 4.4. This is followed by the discussion of our model in comparison to extant models to highlight its key properties. We conclude the section with a description of computational approach.

4.1. Purchase Incidence Model

We model a customer's purchase incidence decisions upon her visits to the online retailer. As discussed, an online customer's tendency to make multiple frequent visits for a purchase decision implies that a purchase conversion may be preceded by a series of clustered visits that occur over a relatively short time period. Within visit clusters, the level of the customer's involvement may increase as she makes more visits for the purchase decision, suggesting that clustered visits could be a precursor of purchase conversion. In addition, the customer's purchase propensity may evolve across visit clusters as she gains more experiences through different shopping episodes. Our model takes into account such effects of clustered visit patterns on purchase behavior. We also account for customer heterogeneity because different customers may have different patterns of store visits that would lead to purchase conversion.

During the period $[0, T]$, where 0 corresponds to the beginning of the data period and T is the censoring point that corresponds to the end of the model calibration period, we observe customer i making J_i visits to the online store at times $t_{i1}, t_{i2}, \ldots, t_{iJ_i}$. We denote customer i's purchase decision at her jth visit as Y_{ij}, where Y_{ij} equals 1 if she purchases and 0 otherwise. We assume that customer i's purchase decision at the jth visit is driven by the latent utility u_{ij}, which has the following relationship with Y_{ij}:

$$Y_{ij} = \begin{cases} 1 & u_{ij} > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

We model u_{ij} as follows:

$$u_{ij} = b_i^u + \boldsymbol{\theta}_i^u \mathbf{V}_{ij} + \vartheta_i^u M_{ij} + \boldsymbol{\beta}^u \mathbf{X}_{ij} + \varepsilon_{ij}. \tag{2}$$

The customer-specific intercept b_i^u captures heterogeneity in customers' tendency to make a purchase. A larger value of b_i^u indicates that the customer's intrinsic purchase propensity is higher than others. The customer-specific coefficients $\boldsymbol{\theta}_i^u$ capture how the individual's purchase propensity changes as a function of her store visit patterns, which are reflected by latent covariates $\mathbf{V}_{ij}$; ϑ_i^u is a customer-specific coefficient for the marketing covariate M_{ij}; and $\boldsymbol{\beta}^u$ is a vector of coefficients that measure the impact of other observed covariates $\mathbf{X}_{ij}$ on the purchase propensity.[1]

With respect to the covariates in Equation (2), $\mathbf{V}_{ij}$ is a vector of covariates which characterize store visit patterns. To consider the impacts of online customers' tendency to make clustered visits on purchase behavior, we include two latent variables, S_{ij} and N_{ij}, constructed based on visit patterns, in $\mathbf{V}_{ij}$. The variable S_{ij} is the number of visit events up to the jth visit made by customer i in a given visit cluster. By definition, S_{ij} is set to 1 at the first visit event of each visit cluster and then increases as the customer makes additional visits within a visit cluster. The variable S_{ij} thus allows us to consider the possibility that the customer's purchase propensity evolves across visits in close temporal proximity as the effects of visit events accumulate within the cluster.[2] The variable N_{ij} is the number of visit clusters up to the jth visit by customer i over the data period. We include N_{ij} to allow for the dynamics that the customer's purchase propensity varies across visit clusters. Importantly, both S_{ij} and N_{ij} can be computed only after the formation of visit clusters is inferred in the visit timing process, which is discussed in Section 4.2. We use log-transformed S_{ij} and N_{ij}, as we expect the diminishing effects of these latent variables on purchase behavior.

The next covariate in Equation (2), M_{ij}, indicates whether or not customer i receives a marketing communication (e.g., e-catalog) from the online retailer between her $(j-1)$th and jth visits.[3] Through another set of observed covariates $\mathbf{X}_{ij}$, we include two relevant factors: (1) whether customer i made a purchase at her prior visit (i.e., $Y_{i,j-1}$), to take into account the dependence of the current purchase decision on the past observation of the outcome variable, and (2) whether the customer's jth visit took place in proximity to major holidays, denoted by $HDAY_{ij}$.[4] The variable $HDAY_{ij}$ is set to 1 if customer i's jth visit occurs during a week prior to a holiday and 0 otherwise.[5]

With the assumption of the type I extreme value distribution for the error term ε_{ij}, the probability that customer i makes a purchase on the jth visit is given by[6]

$$P(Y_{ij}=1\mid t_{ij})=\frac{\exp(b_i^u+\boldsymbol{\theta}_i^u\mathbf{V}_{ij}+\vartheta_i^u M_{ij}+\boldsymbol{\beta}^u\mathbf{X}_{ij})}{1+\exp(b_i^u+\boldsymbol{\theta}_i^u\mathbf{V}_{ij}+\vartheta_i^u M_{ij}+\boldsymbol{\beta}^u\mathbf{X}_{ij})}. \quad (3)$$

4.2. Visit Timing Model

We lay out a model of visit behavior that characterizes store visit patterns of online customers and allows for the specification of S_{ij} and N_{ij} in $\mathbf{V}_{ij}$ in the purchase model. The proposed visit model consists of three main components: an individual-level timing process, the formation of visit clusters, and latent attrition.

4.2.1. Timing Process. We specify the timing model for the set of J_i-1 intervisit times, $t_{i2}-t_{i1}$, $t_{i3}-t_{i2}, \ldots, t_{iJ_i}-t_{i,J_i-1}$, and the right-censored observation $T-t_{iJ_i}$. Following Gupta (1991), who proposed a modeling approach to incorporating time-varying covariates into a multievent timing model, we denote the customer's visit rate at time t between the jth and $(j+1)$th visits as $\lambda_{ij}(t)$ and model the visit rate as

$$\lambda_{ij}(t)=\lambda_i\exp\{\vartheta_i^\lambda M_{ij}(t)+\boldsymbol{\beta}^\lambda\mathbf{Z}_{ij}\}, \quad (4)$$

where λ_i captures heterogeneity across customers in their baseline frequency to visit the online store, and ϑ_i^λ is a customer-specific coefficient for the marketing covariate $M_{ij}(t)$. Customer i may receive a marketing communication at any time, say, t, between her jth and $(j+1)$th visits. The marketing communication could impact the customer's visit rate between t and $t_{i,j+1}$, but not between t_{ij} and t. To reflect this, we let $M_{ij}(t)$ be 0 before a marketing communication is received and 1 afterward. Thus, unlike M_{ij} in Equation (2), the value of $M_{ij}(t)$ may change in the midst of intervisit times, allowing us to account for the timing of the marketing activity. The parameter $\boldsymbol{\beta}^\lambda$ is a vector of coefficients that measure the impact of other observed covariates $\mathbf{Z}_{ij}$ consisting of Y_{ij} and $HDAY_{ij}$.

[1] We allowed $\boldsymbol{\beta}^u$ to be customer specific, but did not find a meaningful improvement in model performance.

[2] We considered the time duration up to the jth visit within a visit cluster in $\mathbf{V}_{ij}$. However, we decided to not include it because of its high correlation with S_{ij}.

[3] We allowed the effect of marketing communication to decay over time (e.g., Ansari et al. 2008). We also tested the model with a stock variable of marketing to consider that marketing communication may have not only immediate impacts but also diminishing carryover effects (e.g., Schweidel and Knox 2013, Schweidel et al. 2014). These alternative specifications of the marketing covariate did not improve the model performance. Additionally, we considered the possibility that the effect of marketing may vary within a visit cluster by including the interaction term of M_{ij} and S_{ij}. The coefficient for the interaction term was estimated to be insignificant.

[4] The major holidays include New Year's Day, Lunar New Year's Day, Children's Day, Independence Day, Thanksgiving Day, and Christmas Day.

[5] We tested the model with the customer's lagged intervisit time and dummy variables for seasonality, and found that the inclusion of the variables only marginally improved the prediction performance of the model.

[6] It is worth discussing that, with a marketing covariate included in the model, the endogeneity issue may arise when firms target specific customers for marketing purposes in a nonrandom manner. The endogeneity of marketing activity is less of a concern in our empirical context, because the online retailer emailed marketing communications to all registered customers in its database, as described in Section 5.1. If marketing activity targeted customers with certain characteristics, one could account for the endogeneity of marketing by modeling the firm's marketing decision and correlating it with the consumer model (e.g., Manchanda et al. 2004).

We assume that the customer's intervisit times follow the Weibull distribution. The Weibull distribution has been widely adopted in marketing literature because of its flexibility and performance (e.g., Jain and Vilcassim 1991, Schweidel et al. 2008). With the time-varying visit rate, the survival function of the Weibull distribution is given by

$$S(t_{i,j+1} \mid t_{ij}) = \exp\left[-\left\{\int_{t_{ij}}^{t_{i,j+1}} \lambda_{ij}(\tau)\,d\tau\right\}^{\nu}\right] \equiv \exp\{-\Lambda(t_{ij}, t_{i,j+1})^{\nu}\}, \tag{5}$$

where ν is the shape parameter of the distribution.

In Equation (5), $\Lambda(t_{ij}, t_{i,j+1}) = \int_{t_{ij}}^{t_{i,j+1}} \lambda_{ij}(\tau)\,d\tau$ can be discretized based on the time period during which the time-varying marketing covariate remains constant (e.g., Gupta 1991, Fader et al. 2004). For example, if the customer has received a marketing communication at time t_{im} in $[t_{ij}, t_{i,j+1}]$, we have $\Lambda(t_{ij}, t_{i,j+1}) = \int_{t_{ij}}^{t_{im}} \lambda_{ij}(\tau)\,d\tau + \int_{t_{im}}^{t_{i,j+1}} \lambda_{ij}(\tau)\,d\tau$. Then, both terms can be easily integrated out because $\lambda_{ij}(\tau)$ remains constant in each integral.

By taking the derivative of the cumulative density function, $F(t_{i,j+1} \mid t_{ij}) = 1 - S(t_{i,j+1} \mid t_{ij})$, we have the probability density function given by

$$f(t_{i,j+1} \mid t_{ij}) = \nu\lambda_{ij}(t_{i,j+1})\Lambda(t_{ij}, t_{i,j+1})^{\nu-1} \cdot \exp\{-\Lambda(t_{ij}, t_{i,j+1})^{\nu}\}. \tag{6}$$

Note that without $M_{ij}(t)$ in Equation (4), we have $\lambda_{ij}(t) = \lambda_{ij}$ for any t in $[t_{ij}, t_{i,j+1}]$ (i.e., the visit rate is time-invariant between the visits), and thus $\Lambda(t_{ij}, t_{i,j+1})$ equals $(t_{i,j+1} - t_{ij})\lambda_{ij}$. Accordingly, $f(t_{i,j+1} \mid t_{ij})$ collapses to the standard density function of the Weibull distribution

$$f(t_{i,j+1} \mid t_{ij}) = \nu\lambda_{ij}\{(t_{i,j+1} - t_{ij})\lambda_{ij}\}^{\nu-1} \cdot \exp[-\{(t_{i,j+1} - t_{ij})\lambda_{ij}\}^{\nu}].$$

It is also worth noting that when ν equals 1 in Equation (6), our Weibull timing process reduces to the exponential timing model with time-varying covariates, proposed by Gupta (1991).

4.2.2. Formation of Visit Clusters. Clustered visit patterns exhibit local concentrations of visit events in close temporal proximity, with each cluster separated by intervening spaces with less dense patterns of visit events. In other words, visit clusters can be represented as a series of visit events with relatively short intervisit times over a short time period, separated by longer intervisit times, then followed by another burst of visit events. This implies that the visit rate within a visit cluster (within-cluster visit rate) is greater than the visit rate between clusters (between-cluster visit rate), which governs intervisit times between the last visit in one visit cluster and the first visit in the subsequent cluster. Moreover, different customers would have different within- and between-cluster visit rates. To incorporate these points into the visit timing process, we specify customer i's baseline visit rate λ_i in Equation (4) as[7]

$$\lambda_i = \begin{cases} \lambda_i^w & \text{customer } i\text{'s within-cluster baseline visit rate,} \\ \lambda_i^b \ (< \lambda_i^w) & \text{customer } i\text{'s between-cluster baseline visit rate.} \end{cases} \tag{7}$$

We also expect that different customers would have different levels of propensity to make clustered visits. Some customers may make more or fewer visits within clusters than others. Furthermore, the visit patterns may change over time for a given customer, as her visit patterns could be more clustered in certain time periods and less clustered in other periods. To account for these dynamics within and between clusters, we consider the transition between the two baseline visit rates and allow the probability of making clustered visits to vary across customers and over time. We denote the probability that customer i stays in the current cluster after the jth visit as p_{ij}. Accordingly, $(1 - p_{ij})$ is the probability that the customer leaves the cluster after the jth visit and forms a new cluster at the $(j+1)$th visit. This indicates that customer i's baseline visit rate between her jth and $(j+1)$th visits is λ_i^w with a probability of p_{ij}, and λ_i^b with a probability of $(1 - p_{ij})$. We specify the clustering probability p_{ij} as

$$p_{ij} = \frac{\exp(b_i^p + \boldsymbol{\theta}_i^p \mathbf{V}_{ij} + \vartheta_i^p M_{ij} + \boldsymbol{\beta}^p \mathbf{Z}_{ij})}{1 + \exp(b_i^p + \boldsymbol{\theta}_i^p \mathbf{V}_{ij} + \vartheta_i^p M_{ij} + \boldsymbol{\beta}^p \mathbf{Z}_{ij})}. \tag{8}$$

The customer-specific intercept b_i^p captures heterogeneity in customers' baseline propensity to make more or fewer visits within a visit cluster. A larger value of b_i^p indicates that the customer is likely to make more visits per cluster than others. The customer-specific coefficients $\boldsymbol{\theta}_i^p$ consider how the individual's propensity to make clustered visits varies depending on the latent covariates S_{ij} and N_{ij} in $\mathbf{V}_{ij}$, which is also included in Equation (2). The inclusion of the covariates allows for the possibility that the clustering probability evolves as the customer makes more visits within a cluster and as she forms more visit clusters over time. The customer-specific coefficient ϑ_i^p captures the impact of the marketing covariate M_{ij} on the clustering probability. The

[7] Our visit timing model can be extended with more than two baseline visit rates, such as a classification of visit clusters with low, middle, and high baseline visit rates. In the preliminary analysis, we found that the benefits of adding more baseline visit rates were marginal. Detailed results are available from the authors on request.

parameter $\boldsymbol{\beta}^p$ is a vector of coefficients for the other observed covariates $\mathbf{Z}_{ij}$.

Equations (6)–(8), taken together, characterize store visit patterns at the individual level and allow us to specify S_{ij} and N_{ij} in $\mathbf{V}_{ij}$. While p_{ij} depends on S_{ij} and N_{ij}, the latent constructs are determined by the clustering probabilities up to the $(j-1)$th visit. The visit rate $\lambda_{ij}(t)$ changes over time, according to the transition between the two baseline visit rates λ_i^w and λ_i^b as well as the observed covariates in Equation (4). Our visit model thereby allows for a number of ways in which visit behavior may vary across customers and over time. For a customer, all else being equal, a larger (smaller) value of λ_i^w leads to smaller (larger) intervisit times within clusters. A larger (smaller) value of p_{ij} indicates that her visit clusters tend to consist of more (fewer) visit events. A larger (smaller) value of λ_i^b indicates smaller (larger) intervisit times between clusters, resulting in more (fewer) visit clusters over a time period for given λ_i^w and p_{ij}.

It is worth discussing the identification of $\boldsymbol{\beta}^\lambda$ and $\boldsymbol{\beta}^p$, coefficient vectors for the observed covariates $\mathbf{Z}_{ij}$ in Equations (4) and (8), respectively. In the visit timing model, Equation (4) specifies the overall visit rate, and Equation (8) describes how the baseline visit rate switches between the within-cluster and between-cluster visit rates. Thus, the effects of $\mathbf{Z}_{ij}$ on the visit rate realize as a binary case only (i.e., λ_i^w or λ_i^b) in Equation (8), whereas the effects of $\mathbf{Z}_{ij}$ are continuous (i.e., $\exp\{\vartheta_i^\lambda M_{ij}(t)+\boldsymbol{\beta}^\lambda \mathbf{Z}_{ij}\}$) in Equation (4). Given these nonlinear specifications, $\boldsymbol{\beta}^\lambda$ and $\boldsymbol{\beta}^p$ can be identified as long as customers have varying sizes of visit clusters and their intervisit times vary within clusters.

4.2.3. Latent Attrition. Because our empirical context is noncontractual in nature, customers may cease shopping at the firm and silently defect at an unobserved time (e.g., Schmittlein et al. 1987). As the failure to account for such latent attrition can result in biased inferences about customer behavior (Fader and Hardie 2009), we incorporate latent attrition by assuming that customer i may become inactive after a visit with a probability of q_i, consistent with the individual-level attrition process of the beta geometric/negative binomial distribution (BG/NBD) model by Fader et al. (2005).[8]

4.3. Incorporating Heterogeneity Across Customers

To account for heterogeneity across customers in our integrated framework of visit and purchase behaviors, we specify the customer-specific model parameters as follows. First, we assume that b_i^u and b_i^p follow a normal distribution. Second, for the two baseline visit rates, we assume that $\lambda_i^w=\lambda_i^b+\delta_i$, where λ_i^b and δ_i follow a lognormal distribution. This ensures that λ_i^w is greater than λ_i^b. Third, we reparameterize the attrition probability q_i as $q_i=\exp(b_i^q)/(1+\exp(b_i^q))$ to ensure that q_i is contained between 0 and 1, and assume that b_i^q follows a normal distribution. We further assume

$$\begin{bmatrix} b_i^u \\ \log\lambda_i^b \\ \log\delta_i \\ b_i^p \\ b_i^q \end{bmatrix} \sim \mathrm{MVN}\left(\begin{bmatrix} \mu_{b^u} \\ \mu_\lambda \\ \mu_\delta \\ \mu_{b^p} \\ \mu_{b^q} \end{bmatrix}, \Sigma\right) \tag{9}$$

to consider the correlation of the model parameters. This specification allows the underlying propensities to visit the store and make a purchase, as well as the tendency to remain in an active relationship, to be associated (e.g., Abe 2009). Last, we assume that the customer-specific coefficients $\boldsymbol{\theta}_i^u$, $\boldsymbol{\theta}_i^p$, ϑ_i^u, ϑ_i^λ, and ϑ_i^p are normally distributed.[9]

4.4. Likelihood Function

Our proposed framework enables us to capture various time-varying patterns of store visits and their association with purchase behavior, while accounting for shopping dynamics within and between visit clusters. However, this benefit comes at a cost. Because we do not observe at which visit event each cluster begins and ends, S_{ij} and N_{ij} in Equations (2) and (8) are not readily available. Without information about the visit clusters, we confront 2^{J_i-1} possible ways of clustering customer i's J_i visits. Moreover, there are three possible scenarios for the right-censored time period, $T-t_{iJ_i}$, depending on whether the customer has dropped out after the J_ith visit and which baseline visit rate drives the censored observation (if the customer remains active after the last visit in the observation period). To deal with this latency embedded in our model, we employ a changepoint modeling framework (e.g., Barry and Hartigan 1993, Fader et al. 2004) in formulating the likelihood function of the model.

The overall likelihood function of customer i's visit and purchase behaviors can be formulated by taking the weighted average of the likelihood function of the observed intervisit times (and the right-censored time period) and purchase conversions associated with each possible formation of visit clusters. We begin by formulating the probability of the cluster formation for the customer. Suppose there are m_i visit clusters for customer i's J_i visits during the data period. We

[8] We allowed the attrition probability to vary across visits depending on the covariates M_{ij} and $\mathbf{Z}_{ij}$. However, we found that none of the covariates had a significant impact on the attrition probability. Moreover, the inclusion of the covariates did not result in improvement in model performance. Accordingly, for the sake of parsimony, we chose not to employ them in specifying the attrition probability.

[9] We allowed these customer-specific parameters to be correlated with others in Equation (9). While this modification significantly increased the computation time for model estimation, it did not result in meaningful findings or improvement in model performance.

denote the latent set of visit clusters as

$$\mathbf{\Gamma}_i = \{\underbrace{(\zeta_{i1}, \xi_{i1})}_{\text{1st cluster}}, \underbrace{(\zeta_{i2}, \xi_{i2})}_{\text{2nd cluster}}, \ldots, \underbrace{(\zeta_{im_i}, \xi_{im_i})}_{m_i\text{th cluster}}\}, \tag{10}$$

where ζ_{ic} and ξ_{ic} indicate the first and last visits in cluster c, respectively. Note that the last cluster may not be fully represented, because data on customer visits are right censored. We thus categorize the last observed visit in the last cluster, denoted by ξ_{im_i}, into the three cases: (1) E1 if customer i becomes inactive after the J_ith visit, (2) E2 if the customer remains active and the last cluster ends with the J_ith visit, and (3) E3 if the customer remains active and the last cluster does not end after the J_ith visit. Because λ_i^w differs from λ_i^b, there are $(2m_i - 2)$ changepoints (at $\xi_{i1}, \zeta_{i2}, \xi_{i2}, \zeta_{i3}, \ldots, \xi_{i,m_i-1}, \zeta_{im_i}$) if $\xi_{im_i} = \text{E3}$; otherwise, there are $(2m_i - 1)$ changepoints (at ξ_{i1}, ζ_{i2}, ξ_{i2}, ζ_{i3}, $\ldots$, ζ_{im_i}, ξ_{im_i}).

Suppose that the 10 visits made by the customer in Figure 1 are partitioned into three visit clusters: the first cluster with the first 3 visits, the second cluster with the next 4 visits, and the third cluster with the last 3 visits. If we also assume that the last visit cluster ends with the last visit and the customer remains active after the last visit to the store, the corresponding cluster formation can be represented as $\mathbf{\Gamma}_i = \{(1,3), (4,7), (8, \text{E2})\}$ with five changepoints. Using p_{ij} and q_i, the probability of the cluster formation for the customer is given by

$$\begin{aligned} P(\mathbf{\Gamma}_i) = {} & \underbrace{\overbrace{(1-q_i)p_{i1}}^{\text{1st visit}} \cdot \overbrace{(1-q_i)p_{i2}}^{\text{2nd visit}} \cdot \overbrace{(1-q_i)(1-p_{i3})}^{\text{3rd visit}}}_{\text{1st cluster}} \\ & \cdot \underbrace{(1-q_i)p_{i4} \cdot (1-q_i)p_{i5} \cdot (1-q_i)p_{i6} \cdot (1-q_i)(1-p_{i7})}_{\text{2nd cluster}} \\ & \cdot \underbrace{(1-q_i)p_{i8}(1-q_i)p_{i9} \cdot (1-q_i)(1-p_{i10})}_{\text{3rd cluster}}, \end{aligned} \tag{11}$$

where, for example, the likelihood of the first cluster is the probability of remaining active after making the first and second visits, multiplied by the probability of leaving the cluster while she remains active after making the third visit. Note that, with the assumed $\mathbf{\Gamma}_i = \{(1,3), (4,7), (8, \text{E2})\}$, the values of S_{ij} and N_{ij} in $\mathbf{V}_{ij}$ become available to specify p_{ij} in Equation (8).

Using the general notation of $\mathbf{\Gamma}_i$ in Equation (10), the probability of the formation of visit clusters $\mathbf{\Gamma}_i$ for customer i is given by

$$\begin{aligned} P(\mathbf{\Gamma}_i) = {} & \prod_{c=1}^{m_i-1} \left\{ \prod_{j=\zeta_{ic}}^{\xi_{ic}-1} (1-q_i)p_{ij} \right\} (1-q_i)(1-p_{i\xi_{ic}}) \\ & \cdot \prod_{j=\zeta_{im_i}}^{J_i-1} (1-q_i)p_{ij} \cdot \{(1-q_i)^{\mathrm{I}(\xi_{im_i} \neq \text{E1})} (1-p_{iJ_i})^{\mathrm{I}(\xi_{im_i} = \text{E2})} \\ & \cdot p_{iJ_i}^{\mathrm{I}(\xi_{im_i} = \text{E3})} + q_i^{\mathrm{I}(\xi_{im_i} = \text{E1})}\}, \end{aligned} \tag{12}$$

where the first-line expression accounts for the likelihood of the first $(m_i - 1)$ visit clusters, and the second-line expression accounts for the likelihood of the last visit cluster and customer attrition, based on the three different scenarios for the right-censored observation.

Our next step is to derive the likelihood functions of the observed intervisit times and purchase conversions, conditional on $\mathbf{\Gamma}_i$. To this end, we denote the set of visit times of customer i and the set of her purchase decisions over the data period as $\mathbf{T}_i = \{t_{i1}, t_{i2}, \ldots, t_{iJ_i}\}$ and $\mathbf{Y}_i = \{Y_{i1}, Y_{i2}, \ldots, Y_{iJ_i}\}$, respectively. The conditional likelihood function of intervisit times is the product of the density and survival functions in Equations (5) and (6), for which the visit rates are determined by $\mathbf{\Gamma}_i$. The conditional likelihood function for the customer in Figure 1 whose 10 visits are partitioned into three visit clusters is given by

$$\begin{aligned} & L(\mathbf{T}_i \mid \mathbf{\Gamma}_i) \\ & = \underbrace{\overbrace{f(t_{i2} \mid t_{i1}; \lambda_i^w)}^{\text{1st int. time}} \overbrace{f(t_{i3} \mid t_{i2}; \lambda_i^w)}^{\text{2nd int. time}}}_{\text{1st cluster}} \overbrace{f(t_{i4} \mid t_{i3}; \lambda_i^b)}^{\text{3rd int. time}} \\ & \quad \cdot \underbrace{f(t_{i5} \mid t_{i4}; \lambda_i^w) f(t_{i6} \mid t_{i5}; \lambda_i^w) f(t_{i7} \mid t_{i6}; \lambda_i^w)}_{\text{2nd cluster}} f(t_{i8} \mid t_{i7}; \lambda_i^b) \\ & \quad \cdot \underbrace{f(t_{i9} \mid t_{i8}; \lambda_i^w) f(t_{i10} \mid t_{i9}; \lambda_i^w)}_{\text{3rd cluster}} S(T \mid t_{i10}; \lambda_i^b), \end{aligned} \tag{13}$$

where, for example, the first two terms in the first-line expression account for the likelihood of the first and second intervisit times within the first cluster, and the last term in the first-line expression accounts for the likelihood of the third intervisit time between the first and second clusters. The survival function in the last line accounts for the right-censored observation. As the likelihoods of the intervisit times and the right-censored observation are conditional on $\mathbf{\Gamma}_i$, the baseline visit rates in the density and survival functions can be determined according to $\mathbf{\Gamma}_i$.

Using the general notation $\mathbf{\Gamma}_i$ in Equation (10), the conditional likelihood function of customer i's intervisit times is given by

$$\begin{aligned} L(\mathbf{T}_i \mid \mathbf{\Gamma}_i) = {} & \prod_{c=1}^{m_i-1} \left\{ \prod_{j=\zeta_{ic}}^{\xi_{ic}-1} f(t_{i,j+1} \mid t_{ij}; \lambda_i^w) \right\} f(t_{i,\xi_{ic}+1} \mid t_{i\xi_{ic}}; \lambda_i^b) \\ & \cdot \prod_{j=\zeta_{im_i}}^{J_i-1} f(t_{i,j+1} \mid t_{ij}; \lambda_i^w) \cdot \{S(T \mid t_{iJ_i}; \lambda_i^b)^{\mathrm{I}(\xi_{im_i} = \text{E2})} \\ & \cdot S(T \mid t_{iJ_i}; \lambda_i^w)^{\mathrm{I}(\xi_{im_i} = \text{E3})}\}, \end{aligned} \tag{14}$$

where the first-line expression accounts for the likelihood of the intervisit times in the first $(m_i - 1)$ visit clusters and the intervisit times between clusters, and the second-line expression accounts for the likelihood of the intervisit times in the last cluster and the right-censored time period.

The derivation of the likelihood function of purchase decisions conditional on $\mathbf{T}_i$ and $\boldsymbol{\Gamma}_i$ is relatively straightforward

$$L(\mathbf{Y}_i \mid \mathbf{T}_i, \boldsymbol{\Gamma}_i) = \prod_{j=1}^{J_i} P(Y_{ij} = 1 \mid t_{ij})^{\mathrm{I}[Y_{ij}=1]} \cdot \{1 - P(Y_{ij} = 1 \mid t_{ij})\}^{\mathrm{I}[Y_{ij}=0]}, \quad (15)$$

where $\mathrm{I}[\cdot]$ is an indicator function in which $\mathrm{I}[\cdot]$ is 1 if $[\cdot]$ is true and 0 otherwise. Conditional on $\boldsymbol{\Gamma}_i$, the values of S_{ij} and N_{ij} in $\mathbf{V}_{ij}$ become available to specify $P(Y_{ij} = 1 \mid t_{ij})$ in Equation (3).

The likelihood function in Equation (14) is conditional on $\boldsymbol{\Gamma}_i$, the likelihood function in Equation (15) is conditional on $\mathbf{T}_i$ and $\boldsymbol{\Gamma}_i$, and Equation (12) is the probability that the cluster formation is given by $\boldsymbol{\Gamma}_i$. To derive the unconditional likelihood function for customer i, applying the Bayes' theorem, we weight Equation (15) by Equation (14) and then by Equation (12) over all possible $\boldsymbol{\Gamma}_i$'s

$$L(\mathbf{T}_i, \mathbf{Y}_i) = \sum L(\mathbf{Y}_i \mid \mathbf{T}_i, \boldsymbol{\Gamma}_i) L(\mathbf{T}_i \mid \boldsymbol{\Gamma}_i) P(\boldsymbol{\Gamma}_i), \quad (16)$$

where the summation encompasses the $2^{J_i-1} \cdot 3$ possible ways of clustering the customer's visits with the consideration of latent attrition.

Finally, we obtain the overall likelihood function of the model by incorporating customer heterogeneity into Equation (16)

$$L = \prod_{i=1}^{I} \left[\int L(\mathbf{T}_i, \mathbf{Y}_i)\, dF(b_i^u, \log \lambda_i^b, \log \delta_i, b_i^p, b_i^q, \boldsymbol{\theta}_i^u, \boldsymbol{\theta}_i^p, \vartheta_i^u, \vartheta_i^\lambda, \vartheta_i^p) \right], \quad (17)$$

where $dF(\cdot)$ denotes the joint probability density function for the customer-specific parameters.

4.5. Discussion of the Model

The proposed model infers a customer's visit patterns and links them to her purchase behavior in a noncontractual setting. In doing so, our model generalizes extant multievent timing models in customer base analysis (e.g., Schmittlein et al. 1987, Fader et al. 2005, Abe 2009) by allowing for the clustered visit patterns of online customers and shopping dynamics within and between clusters. We compare and contrast our model with several benchmark models to highlight its key properties.

In its basic framework, our visit timing model is built on the BG/NBD model, an established "buy 'til you die" model proposed by Fader et al. (2005). The BG/NBD model assumes a stationary exponential timing process with a time-invariant arrival rate until the customer becomes inactive. As a result, the model has a limited ability to capture the clustered patterns of store visits. Our model extends the BG/NBD model by employing a more flexible underlying distribution (i.e., Weibull) for intervisit times and allowing for the time-varying transition between the two baseline visit rates, λ_i^w and λ_i^b. If the covariates $M_{ij}(t)$ and $\mathbf{Z}_{ij}$ are ignored in Equation (4) and the transition between the baseline visit rates is not considered (i.e., $p_{ij} = 0$), the essence of our visit model reduces to the BG/NBD model when ν is restricted to 1.

Our proposed changepoint model can be regarded as a generalization of the HMM framework. In its standard form, a HMM is a finite-state stochastic model in which the observed behavioral outcome at time t depends on the latent state of the system at the time, and the transitions between states occur according to the Markov process. HMMs have been employed to model a wide range of latent changes in consumer behavior (e.g., Montgomery et al. 2004, Netzer et al. 2008, Schweidel et al. 2009). Unlike HMMs, in our model, the transition process depends on latent states (namely, "alive within a visit cluster," "alive between clusters," and "dead") of multiple periods back, and the order of dependence varies over time, because the clustering probability changes with S_{ij} and N_{ij} within and between clusters. If the high-order dependence on the latent states is ignored, our model collapses down to a three-state HMM.

From a modeling standpoint, our model shares similarities with the changepoint model proposed by Fader et al. (2004) for new product sales forecasting. Both models update a customer's visit and/or purchase rates autonomously to capture underlying nonstationarity in the customer arrival processes. However, an important difference pertains to how the models account for the probability of a changepoint occurring in the timing process, a key component of the changepoint model. Fader et al. (2004) specify the probability that a customer updates her purchase rate at the aggregate level and assume that this probability decreases monotonically as the customer gains experience with a new product.[10] By contrast, we model p_{ij} at the individual level and allow the probability to increase or decrease over time, depending on the past visit patterns and observed covariates.[11]

[10] We acknowledge that this assumption is reasonable for Fader et al. (2004) because customers would become less likely to change their preferences for the new product as they gain more experience with it.

[11] Another notable difference is that Fader et al. (2004) allow the purchase rate to be updated to any nonnegative value at each changepoint without the explicit consideration of latent customer attrition. By comparison, the visit rate in our model switches between two values while the customer remains active.

4.6. Computational Approach

We adopt a Bayesian approach and use the Markov Chain Monte Carlo (MCMC) methods to estimate the proposed model. Central to our Bayesian estimation approach is to treat S_{ij} and N_{ij} as latent variables and draw their samples through data augmentation (Tanner and Wong 1987). Conditional on S_{ij} and N_{ij}, Equations (3) and (8) are nothing but a typical logistic function, and thus the estimation of other model parameters becomes straightforward. However, because S_{ij} and N_{ij} can take any positive integer values, we find it difficult to obtain the sample draws of the variables directly. To deal with this, we define a latent variable K_{ij} that indicates whether customer i stays within the current cluster after the jth visit ($K_{ij}=1$) or leaves the cluster after the visit ($K_{ij}=0$). Then, by the definition of p_{ij}, K_{ij} equals 1 with a probability of p_{ij} and 0 with a probability of $(1-p_{ij})$. Once we draw samples of K_{ij} on the basis of customers' shopping patterns, S_{ij} and N_{ij} can be obtained from the following relationships with K_{ij}:

$$S_{ij} = S_{i,\,j-1} \times K_{i,\,j-1} + 1 \quad \text{and} \quad N_{ij} = \sum_{m=1}^{j-1} K_{im} + 1. \tag{18}$$

Conditional on S_{ij} and N_{ij} at each MCMC iteration, other parameters can be drawn from their respective full conditional distributions derived using the standard Bayesian theory.

To complete the Bayesian specification of the model, we assume noninformative conjugate priors to the model parameters. For aggregate-level parameters and mean parameters, we use a normal density prior. For the variance–covariance matrix, we assume that the inverse of the matrix follows a Wishart distribution. The model is then estimated using the Gibbs sampler with the Metropolis–Hastings steps. Implementation of this method is relatively easy in WinBUGS, a publicly available software package, once the data augmentation scheme is established. We believe the use of WinBUGS can help facilitate the application and replication of our model by researchers and practitioners who utilize the software. Sample draws obtained from the MCMC algorithm are used to compute summary measures of the parameter estimates. Results reflect the output of MCMC draws for 30,000 iterations, with a burn-in period of 30,000 iterations.

5. An Empirical Application

In this section, we provide an empirical demonstration of our proposed model. We begin by describing the data used in the analysis. We next discuss the performance of our model in comparison to other benchmark models and demonstrate the model fit results. Finally, we describe inferences based on parameter estimates and discuss managerial implications of model results.

Table 1 Descriptive Statistics

	Mean	Std. dev.	Min.	Max.
Intervisit time (days)	17.8	34.3	0.0	350.3
No. of visits	13.3	9.5	4.0	68.0
No. of purchases	1.5	1.8	0.0	12.0

5.1. Data

Data for our empirical application were provided by an online fashion retailer in Korea. The data contain information on the visit (Internet session) and purchase records of online shoppers at the website for a period of 16 months from February 2013 to May 2014. The data include the date and time of each online visit, allowing us to compute the precise intervisit times for individual customers, and whether the customers made a purchase during their visit to the e-commerce site. The data also contain information on the online retailer's marketing efforts, in which it sent customers a series of email communications for advertising and promotional purposes. The contents of emails vary across occasions, including e-catalogs, promotions, and new product arrivals. The retailer sent marketing communications seven times over the data period.[12] Each of the emails was sent to all customers registered in the retailer's customer database.

We use the first 12 months of data for model calibration and the remaining 4 months of data for model validation. When a customer makes only a few visits, her visit patterns are less likely to be identified with statistical significance. We thus restrict our attention to customers who made more than three visits (i.e., two intervisit times) to the online store over the 12-month calibration period.[13] We randomly sample 1,000 customers for our analysis. Table 1 summarizes the descriptive statistics of the calibration data. On average, customers made 13.3 visits to the online store, with a mean intervisit time of 17.8 days. Out of all visits, 11.2% resulted in purchase conversion.

As our model utilizes the clustered patterns of customer visits in investigating online shopping behavior, it is desirable to check whether our data actually exhibit clustered visit patterns before we apply the proposed model to the data. One may consider visually examining individual customers' visit patterns. The approach, however, becomes very burdensome with a large number of customers. Furthermore, when data patterns are time varying with occasional occurrences of clustered visits, it could be difficult to conclude the existence and the extent of the clustered

[12] The marketing communications were delivered on May 14, 2013; July 5, 2013; September 8, 2013; November 24, 2013; December 22, 2013; January 23, 2014; and March 18, 2014.

[13] In our data, 68% of customers made more than three visits to the online store, and they account for 93% of total purchase transactions.

patterns. Alternatively, one may consider drawing the histogram of intervisit times at the customer level with an expectation that a high frequency of short intervisit times indicates clustered visit patterns. Such observation is, however, a necessary condition rather than a sufficient one for clustered patterns, as it can also arise in random event patterns. More important, these descriptive approaches do not provide any measures that quantify the clumpiness of data, which makes it difficult to conclude whether and how much the overall patterns are clustered.

The needs for a well-defined measure of data clumpiness have triggered a stream of research. Recently, Zhang et al. (2013) proposed a new class of clumpiness measures that are shown to have high statistical power and robust performance. We employ their entropy metric H_p to measure the clumpiness of individuals' visit patterns in our data.[14] The mean clumpiness across customers is 0.36, with a standard deviation of 0.18. Following Zhang et al. (2013), we test the significance of the clumpiness value for each customer, based on the Z-table, the table of clumpiness critical values, generated via Monte Carlo simulations. In the test, the null hypothesis of randomness is rejected and the customer is judged to have clustered patterns at an α significance level when the clumpiness measure is larger than the corresponding α-level clumpiness critical value. In our data, 71% of customers are judged to have clustered visit patterns at a 0.05 significance level, suggesting that this is a widespread phenomenon across online customers.

We next explore the association of purchase behavior with store visit patterns. As a simple means to consider clustered visit patterns, we categorize each customer's visit events into two groups, based on whether the preceding intervisit time is smaller than the median intervisit time of the customer. We then compute the conversion rate for each group of her visits and take an average of each set of conversion rates across customers. We find that the conversion rate at visits with intervisit times larger than the customer-specific median is 0.088. By comparison, the conversion rate at the visits with intervisit times less than the customer-specific median is much higher at 0.125. These summary statistics suggest the possibility that clustered visit activities are associated with purchase conversions, implying that taking into account clustered visit patterns can help us better understand online customer behavior.

5.2. Model Comparison

We estimate a number of benchmark models to assess the importance of considering clustered visit patterns in predicting online shopping behavior. Accordingly, the alternative models we examine vary with respect to whether and how to account for the clustering phenomenon and its impact on purchase behavior.

The first two benchmark models (Models 1 and 2) serve as base models, ignoring the notion of clustering by assuming a single baseline visit rate that drives the visit timing process. Model 1 assumes that the visit rate is specific to individual customers but constant over time. This model does not include any covariates in Equation (4) and ignores the transition between the two baseline visit rates λ_i^w and λ_i^b by assuming $p_{ij} = 0$ in Equation (8). It also assumes the purchase probability to be time-invariant with no covariates in Equation (2). Model 2 extends Model 1 by allowing the purchase probability and the visit rate to vary over time, depending on the sets of observed covariates, M_{ij} and $\mathbf{X}_{ij}$ in Equation (2) and $M_{ij}(t)$ and $\mathbf{Z}_{ij}$ in Equation (4). Yet, Model 2 still fails to consider the clustered patterns of store visits.

The second set of the benchmark models (Models 3 and 4) allows for clustered visit patterns and their impact on purchase behavior, but employs deterministic approaches to clustering. Model 3 is formulated on the basis of the notion that a purchase conversion may indicate the formation of visit clusters. This model assumes that a visit cluster spans all visits between two consecutive purchase transactions, regardless of their temporal proximity. Thus, the effects of visits accumulate until a purchase event occurs, at which time the net visit effect resets to zero (e.g., Moe and Fader 2004). With the deterministic clustering, the values of S_{ij} and N_{ij} in $\mathbf{V}_{ij}$ become observed, and our dynamic changepoint model reduces to a mixture of binary logit and Weibull timing processes. In Model 4, we compute the median of each customer's intervisit times and use the value to cluster her visits with an assumption that any consecutive visits with intervisit times less than the median constitute a visit cluster.[15] As a result, this deterministic approach to clustering ignores the possibility that all visits by an individual constitute one cluster, or each visit forms a cluster.

There are four benchmark models under our proposed modeling framework. Model 5 considers clustered visit patterns by allowing for different baseline visit rates for within- and between-cluster visits in a probabilistic manner. However, it does not include any covariates in Equations (2), (4), and (8). Model 6 extends Model 5 by including observed covariates, M_{ij}, $\mathbf{X}_{ij}$, $M_{ij}(t)$, and $\mathbf{Z}_{ij}$, in the equations. Yet, it still ignores latent covariates $\mathbf{V}_{ij}$ in Equations (2) and (8) and thus fails to consider the impacts of clustered

[14] We refer readers to Zhang et al. (2013) for the specification of the measure and detailed discussions on it.

[15] We thank an anonymous reviewer for suggesting this benchmark model.

Table 2 **Model Comparison**

	LMD	DIC	In-sample hit rate		Out-of-sample hit rate	
			Visit	Purchase	Visit	Purchase
No clustering						
Model 1	−27,144	54,498	0.24	0.65	0.22	0.62
Model 2	−26,237	52,692	0.27	0.81	0.24	0.77
Deterministic clustering						
Model 3	−24,565	49,375	0.34	0.80	0.31	0.76
Model 4	−22,712	45,670	0.47	0.82	0.43	0.79
Latent clustering						
Model 5	−20,985	42,212	0.51	0.65	0.46	0.62
Model 6	−20,426	41,103	0.53	0.81	0.48	0.77
Model 7	−20,245	40,750	0.54	0.82	0.48	0.79
Model 8	−20,102	40,656	0.54	0.83	0.49	0.80

visit patterns on customer behavior. Model 7 is built on a fully parameterized version of our proposed model. However, instead of S_{ij} and N_{ij} in $\mathbf{V}_{ij}$, which are inferred in the proposed model, it includes two observed covariates constructed based on a customer's visit events: (1) the number of visits made on the same day with the jth visit and (2) the number of days on which the customer made at least one visit to the online store. These observed variables are employed to mimic the roles of S_{ij} and N_{ij} in our full model. Finally, Model 8 is our proposed model.

To compare the performance of the alternative models, we compute the log marginal density (LMD; Newton and Raftery 1994) and the deviance information criterion (DIC; Spiegelhalter et al. 2002) of the models. As several researchers caution against solely relying on the likelihood-based fit measures because of their potential instability (e.g., Gelman and Rubin 1995, Schwartz et al. 2014), we also perform one-step-ahead predictions in which we forecast a customer's next intervisit time and purchase decision. For visit behavior, we forecast a customer's intervisit time between her jth and $(j+1)$th visits conditional on her visit and purchase events up to the jth visit. We then compute the hit rate, which refers to the percentage of times that the model predicts the next visit time with an error of less than one day. For purchase behavior, we forecast a customer's purchase decision at her $(j+1)$th visit conditional on her past shopping history and compute the hit rate on whether the model correctly predicts the purchase decision. The hit rates are computed for each individual and averaged across customers for reporting.

Table 2 reports the model fit measures. We find that the model performance substantially improves as we allow for clustered visit patterns in the visit timing process (Models 1 and 2 versus Models 3–8) and a latent clustering of store visits in a probabilistic approach (Models 3 and 4 with deterministic clustering versus Models 5–8 with latent clustering). Between the two base models with no clustering, Model 2 provides a better fit than Model 1, implying that the observed covariates considered in our study are useful in predicting online shopping behavior. Comparing the base models and the benchmark models with deterministic clustering, we find that it is useful to account for clustered visit patterns even through the deterministic approaches. The better fit of Model 8 over Model 3 implies that purchases may not be necessarily associated with visit clusters, perhaps because some customers may make no purchases after making multiple frequent visits to the store. The better performance of Model 8 over Model 4 also indicates the benefits of considering visit patterns in a probabilistic manner. Overall, the results demonstrate the efficacy of our proposed approach to modeling online shopping patterns, compared to others with no clustering or deterministic clustering.

When it comes to the comparison among benchmark models with latent clustering, Model 6 outperforms Model 5 due to the inclusion of the observed covariates. We find that Model 8 provides a better fit than Model 6, suggesting that taking into account the effect of visit patterns on shopping dynamics can help better predict online customer behavior. Model 8 outperforms Model 7, although the difference in the fit measures between the models is relatively small. This result may imply that there are visit clusters that span more than a day, possibly because some customers make clustered visits to the online store over a longer period. Therefore, simply counting the number of daily visits cannot fully capture the impact of clustered visits on customer behavior.

To further investigate the benefits of considering clustered visit patterns in predicting online purchase behavior, we look at confusion matrices for Models 6 and 8, generated based on the observed and predicted outcomes of purchase decisions in the one-step-ahead forecasting. Tables 3 and 4 present the confusion matrices in which the cell entries indicate

Table 3 Confusion Matrix for Model 6

	Predicted		
	Purchase	No purchase	Sum
Observed			
Purchase	254	1,244	1,498
No purchase	1,252	10,574	11,826
Sum	1,506	11,818	13,324

Table 4 Confusion Matrix for Model 8

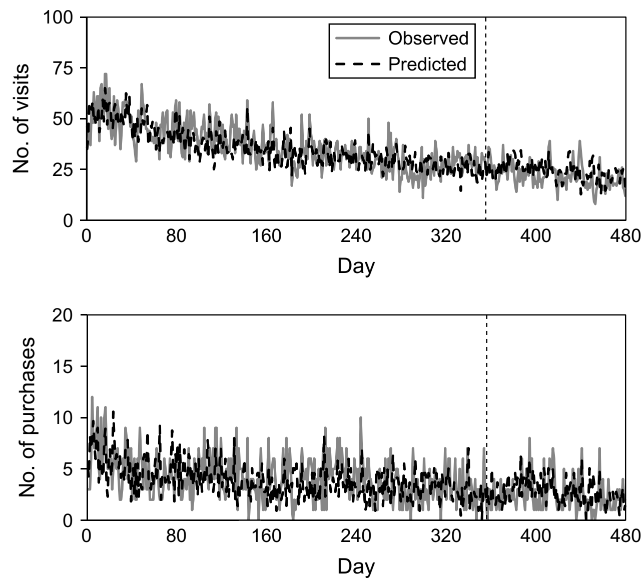
	Predicted		
	Purchase	No purchase	Sum
Observed			
Purchase	288	1,210	1,498
No purchase	1,099	10,727	11,826
Sum	1,387	11,937	13,324

the number of cases for the corresponding conditions. Using the tables, we calculate the hit rates separately for visits at which a purchase occurred (i.e., the true positive hit rate) and for visits at which no purchases occurred (i.e., the true negative hit rate). The true positive hit rate is 0.17 (= 254/1,498) for Model 6 and 0.19 (= 288/1,498) for Model 8, and the true negative hit rate is 0.89 (= 10,574/11,826) for Model 6 and 0.91 (= 10,727/11,826) for Model 8. Thus, Model 8 outperforms Model 6 for both cases. In particular, for visits with purchases, the true positive hit rate of Model 8 is 13% higher than that of Model 6. This suggests that taking into account the association of purchase behavior with clustered visit patterns can indeed improve the predictions of customers' visit-to-purchase conversion behavior.

5.3. Model Validation

To validate the performance of our model, we forecast customers' shopping behavior across store visits. The predictions are performed by fully simulating all outcome measures over the entire data period. Figure 3 compares the predicted posterior means of the number of daily visits and purchases to the corresponding observed values, aggregated across customers. As shown, the model closely captures the overall patterns of daily customer behavior. Using these results, we depict the cumulative number of visits and purchases by the customers in Figure 4. At the end of the calibration (validation) period, the model predicts the cumulative visits at a 1.3% (1.6%) error rate and the cumulative purchases at a 0.9% (1.8%) error rate, indicating that the model can accurately track customers' visit and purchase frequencies. As another means to assess our model, we compare the distribution of the observed and predicted intervisit times in Figure 5, and the distribution of the observed and

Figure 3 Model Prediction on Daily Number of Visits and Purchases



predicted total number of visits and purchases across customers in Figure 6. These results demonstrate the capability of our model to predict online customer behavior.

We also examine whether our model properly reflects the clustered patterns of online store visits. In Figure 7, we select a sample of customers with different visit patterns and show their observed and predicted visit data. As a means to assess the model's ability to capture clustered visit patterns, we compute the clumpiness measures proposed by Zhang et al. (2013) for individual customers' observed and predicted visit data separately. We then correlate these measures (i.e., one with observed data and the

Figure 4 Model Prediction on Cumulative Number of Visits and Purchases

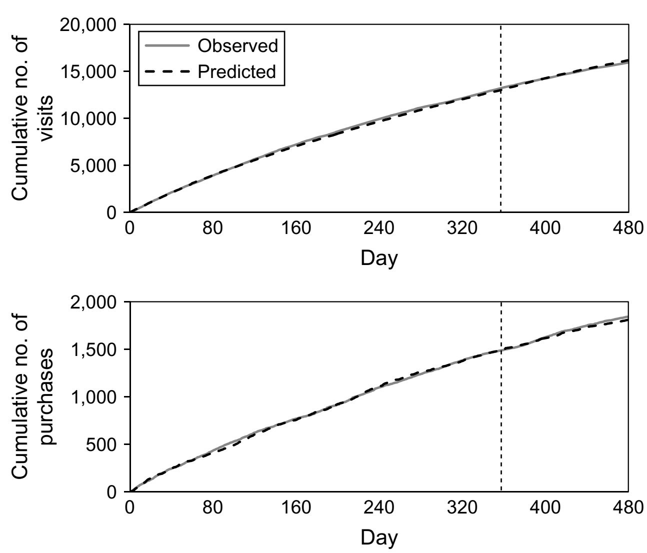
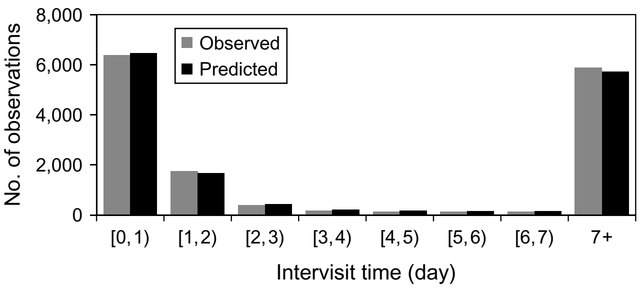


Figure 5 Model Prediction on the Distribution of Intervisit Times



other with predicted data) across customers. We find that the correlation coefficient is 0.47, indicating the model's ability to reflect the clustering phenomenon. By comparison, when the predictions are made by the no-clustering models that do not account for clustered visit patterns (i.e., Models 1 and 2 in Table 2), the correlation of the clumpiness measures is 0.09 only.

5.4. Parameter Inferences

We describe inferences based on the estimates of our model parameters. Tables 5 and 6 report the posterior means and 95% posterior intervals for the parameter estimates of the purchase and visit models, respectively.

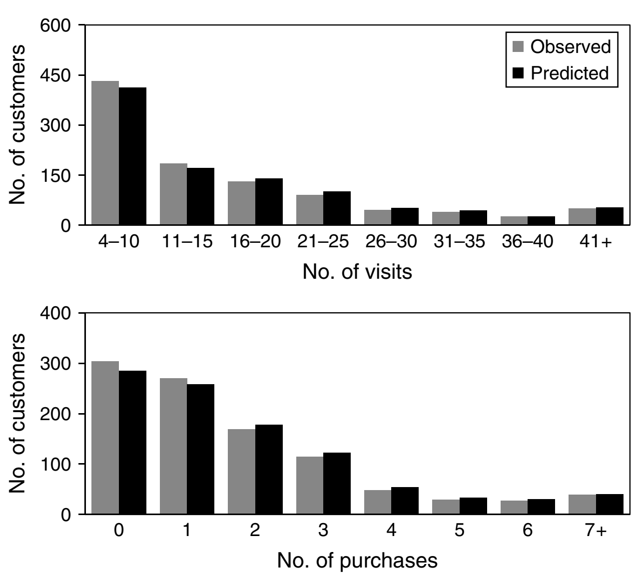
5.4.1. Effect of Visit Patterns. From the estimate of $\mu_{\theta_1^u}$ in Table 5, which is the mean parameter of the individual-level coefficients for S_{ij} in the purchase model, we find that the number of store visits within a cluster is positively associated with the purchase propensity. This implies that the effects of visits on purchasing tend to accumulate within a visit cluster. By comparison, the estimate of $\mu_{\theta_2^u}$, which is the mean parameter of the individual-level coefficients for N_{ij}, suggests that the purchase probability does not correlate with the number of visit clusters by customers in the past. This indicates that in our data, customers' overall purchase propensity does not evolve directionally across different shopping episodes.

Given the significant impact of S_{ij} on the purchase propensity, we examine how the purchase probability changes within visit clusters. Using the cluster formation inferred from the estimates of K_{ij}'s, we classify visit clusters with respect to the number of visits that constitute the clusters. We then compute the conversion rates (i.e., conditional purchase probabilities) across visits within clusters of the same size. We find that, on average, customers make 2.18 visits per cluster, with a standard deviation of 1.96, and the conversion rates vary substantially depending on both the size of a visit cluster and the location of a visit event in the cluster. When a cluster consists of only one visit (47% of visit clusters are one-visit clusters), the conversion rate is 0.10. Figure 8 shows how the conversion rates vary within clusters when a cluster consists of more than one visit. For visit clusters with two visits, which account for 25% of visit clusters, the conversion rate ranges from 0.08 at the first visit to 0.12 at the second visit. For visit clusters with three visits, which account for 13% of visit clusters, the conversion rate is 0.08 at the first visit, increases to 0.09 at the second visit, and then rises to 0.13 at the third visit. Overall, the purchase likelihoods tend to be smaller at earlier visits and higher at later visits within clusters.[16]

From the estimates of customer-specific parameters, $\log \lambda_i^b$ and $\log \delta_i$, whose posterior mean and variance are reported in Table 6, we find that the mean between-cluster baseline visit rate (λ_i^b) is 0.02 and the mean within-cluster baseline visit rate (λ_i^w) is 1.53. We then compute the mean within-cluster and between-cluster intervisit times, based on the inferred formation of visit clusters. From the results, we find that the mean between-cluster intervisit time is 43.68 days. In sharp contrast, the mean within-cluster intervisit time is only 0.81 days. Customers therefore make considerably more frequent visits within clusters. This substantial difference between the within- and between-cluster intervisit times plays a critical role in predicting customers' visit timing behavior, as shown in Section 5.2.

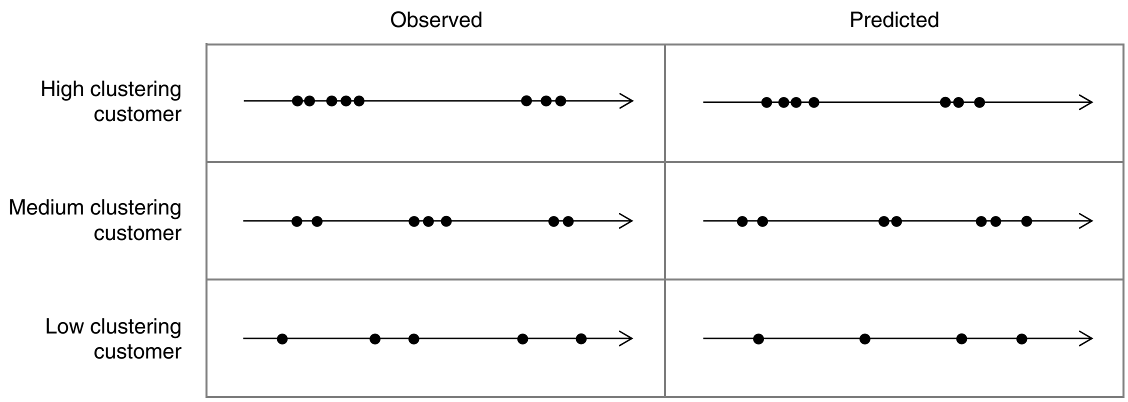
When it comes to the effect of the latent variables S_{ij} and N_{ij} on the clustering probability, the estimate

Figure 6 Model Prediction on the Distribution of Number of Visits and Purchases



[16] We do not, however, find such patterns of increasing conversion rates for clusters with more than seven visits. Perhaps this could be because a sizable portion of these large clusters are formed by knowledge-building shoppers (Moe 2003) who visit the store to learn and process informational content without the intention of buying.

Figure 7 Model Prediction for Selected Customers



of $\mu_{\theta_1^p}$, which is the mean parameter of the individual-level coefficients for S_{ij}, indicates that a customer's likelihood of making additional visits within a cluster decreases as the cluster size increases. By comparison, from the estimate of $\mu_{\theta_2^p}$, which is the mean parameter of the individual-level coefficients for N_{ij}, we find that the number of visit clusters formed in the past does not significantly affect the clustering probability for future visits.

5.4.2. Effect of Observed Covariates. The next set of results pertains to the effects of observed covariates on online shopping behavior. From the estimate of μ_{ϑ^u} in Table 5, which is the mean parameter of the individual-level coefficients for M_{ij} in the purchase model, we find that the firm's email communication to customers increases their purchase propensity. The estimates of μ_{ϑ^λ} and μ_{ϑ^p} in Table 6, which are the mean parameters for marketing covariates in the visit model, suggest that the marketing activity reduces customers' intervisit times and increases their tendency to make clustered visits.

As the firm's email communication significantly affects visit and purchase behaviors, we quantify its impacts by simulating customer behavior with and without the presence of marketing in our model. Assuming the mean values for other covariates, we find that on average an email communication reduces a customer's current intervisit time by 6%, from 0.87 days to 0.82 days within a visit cluster and from 43.05 days to 40.58 days between clusters. Upon the customer's visit that follows the reception of the marketing message, the purchase probability increases by 30%, from 0.10 to 0.13, and the clustering probability increases by 19%, from 0.52 to 0.61. Using the cluster formation inferred from the estimates of K_{ij}'s, we then look into the impacts of the marketing activity on customer behavior within clusters. The results reveal that

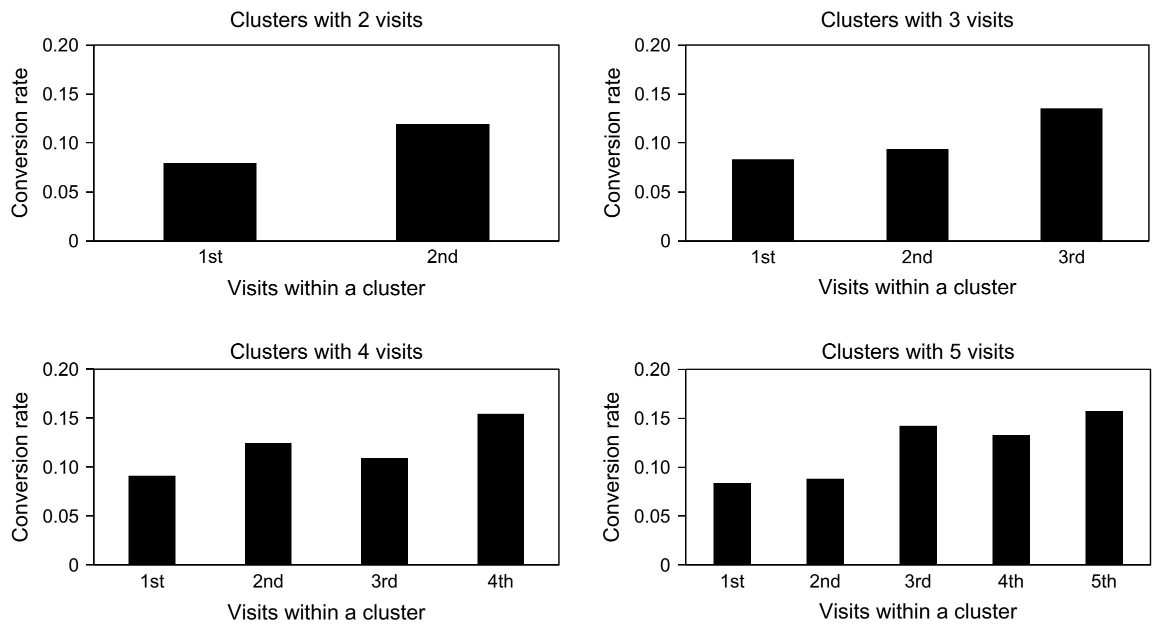
Table 5 Parameter Estimates of the Purchase Model

Parameter	Description	Posterior mean	95% posterior interval
μ_{b^u}	Intercept: mean	−2.547	[−2.662, −2.444]
$\sigma^2_{b^u}$	Intercept: variance	0.641	[0.491, 0.801]
$\mu_{\theta_1^u}$	S_{ij}: mean	0.369	[0.240, 0.495]
$\sigma^2_{\theta_1^u}$	S_{ij}: variance	0.165	[0.073, 0.271]
$\mu_{\theta_2^u}$	N_{ij}: mean	0.056	[−0.011, 0.128]
$\sigma^2_{\theta_2^u}$	N_{ij}: variance	0.030	[0.011, 0.055]
μ_{ϑ^u}	Marketing: mean	0.310	[0.121, 0.508]
$\sigma^2_{\vartheta^u}$	Marketing: variance	0.273	[0.007, 0.596]
$\mathbf{X}_{ij}$			
$Y_{i,j-1}$	Lagged purchase	−0.678	[−0.886, −0.477]
$HDAY_{ij}$	Holiday	0.141	[−0.082, 0.352]

Table 6 Parameter Estimates of the Visit Model

Parameter	Description	Posterior mean	95% posterior interval
ν	Shape parameter	2.229	[2.192, 2.265]
λ_{ij}			
μ_λ	Baseline λ_i^b: mean	−4.113	[−4.186, −4.051]
σ^2_λ	Baseline λ_i^b: variance	0.837	[0.758, 0.923]
μ_δ	Baseline δ_i: mean	0.267	[0.227, 0.302]
σ^2_δ	Baseline δ_i: variance	0.309	[0.280, 0.342]
μ_{ϑ^λ}	Marketing: mean	0.054	[0.014, 0.097]
$\sigma^2_{\vartheta^\lambda}$	Marketing: variance	0.007	[0.001, 0.017]
$\mathbf{Z}_{ij}$			
Y_{ij}	Purchase	−0.594	[−0.623, −0.564]
$HDAY_{ij}$	Holiday	0.290	[0.253, 0.326]
p_{ij}			
μ_{b^p}	Intercept: mean	0.309	[0.223, 0.401]
$\sigma^2_{b^p}$	Intercept: variance	0.570	[0.471, 0.673]
$\mu_{\theta_1^p}$	S_{ij}: mean	−0.114	[−0.185, −0.047]
$\sigma^2_{\theta_1^p}$	S_{ij}: variance	0.016	[0.004, 0.063]
$\mu_{\theta_2^p}$	N_{ij}: mean	−0.047	[−0.100, 0.004]
$\sigma^2_{\theta_2^p}$	N_{ij}: variance	0.014	[0.002, 0.028]
μ_{ϑ^p}	Marketing: mean	0.435	[0.245, 0.638]
$\sigma^2_{\vartheta^p}$	Marketing: variance	0.032	[0.002, 0.069]
$\mathbf{Z}_{ij}$			
Y_{ij}	Purchase	−0.346	[−0.488, −0.216]
$HDAY_{ij}$	Holiday	0.011	[−0.144, 0.165]
q_i			
μ_{b^q}	Intercept: mean	−3.914	[−3.942, −3.890]
$\sigma^2_{b^q}$	Intercept: variance	0.126	[0.111, 0.142]

Figure 8 **Conversion Rates Within a Visit Cluster**



when a visit cluster follows an email communication or contains it within the cluster, the size of the cluster increases by 23%, from 2.01 visits to 2.48 visits, and the conversion rate per cluster increases by 41%, from 0.17 to 0.24.

Looking at the estimates of the coefficients for $\mathbf{X}_{ij}$ in Table 5, we find that $Y_{i,j-1}$ is negatively associated with the purchase propensity. Thus, customers are less likely to make a purchase following their purchase at a prior visit. The effect of $HDAY_{ij}$ on the purchase probability is estimated to be insignificant. With respect to the estimates of the coefficients for $\mathbf{Z}_{ij}$ in Table 6, we find that Y_{ij} has a significantly negative impact on λ_{ij} and p_{ij}. This indicates that when a customer makes a purchase at her prior visit, the time until the next visit gets longer, and the customer is more likely to initiate a new cluster at her next visit. Thus, customers' purchase history is closely tied to their visit behavior, as expected from the multivisit tendency for a purchase decision online. We find that $HDAY_{ij}$ is positively associated with λ_{ij}, which implies that customers tend to make more frequent visits during holidays.

5.4.3. Customer Attrition and Correlation Structure. We report the model results on customer attrition and the variance–covariance matrix. From the estimate of the customer-specific parameter b_i^q, we find that, on average, customers defect with a probability of 0.02 after each visit to the online store. Thus, at the end of the calibration period, an average customer is active with a probability of 0.76 after making 13.3 visits to the store.

Table 7 reports the estimated variance–covariance matrix Σ. The variance estimates show that customers are heterogeneous in their visit and purchase behaviors. The covariance estimates suggest that it is important to consider interdependence between the model components within customers. For example, the negative covariance between b_i^u and $\log\lambda_i^b$ indicates that customers who exhibit a longer intervisit time between clusters tend to have a higher baseline purchase propensity. The positive covariance between $\log\lambda_i^b$ and $\log\delta_i$ indicates that the within-cluster and between-cluster baseline visit rates are positively associated within individuals.

5.5. Individual-Level Store Visit Patterns

Our proposed model and Bayesian estimation approach allow us to obtain the estimates of customer-specific parameters, which provide a set of useful inferences regarding store visit patterns at the individual customer level. Table 8 reports the posterior means of several customer-specific statistics, computed from the estimates of individual-level parameters across the iterations of the MCMC samplers: (1) the number of visits per cluster, (2) the intervisit times within a cluster, (3) the time length of a cluster, (4) the intervisit times between clusters, and (5) the number of clusters during the data period, at the individual customer level. For the purpose of illustration, we show the top half of the customers at every fifth percentile with respect to their total number of visits during the data period, because there are relatively small variations in these statistics for the customers with fewer visits. The customer-level

Table 7 Estimated Σ

	b_i^u	$\log\lambda_i^b$	$\log\delta_i$	b_i^p	b_i^q
b_i^u	0.641 [0.491, 0.801]	−0.267 [−0.344, −0.180]	−0.185 [−0.233, 0.136]	0.184 [0.076, 0.310]	−0.140 [−0.180, −0.104]
$\log\lambda_i^b$		0.837 [0.758, 0.923]	0.379 [0.338, 0.424]	−0.323 [−0.393, −0.256]	0.257 [0.226, 0.289]
$\log\delta_i$			0.309 [0.280, 0.342]	−0.243 [−0.293, −0.196]	0.119 [0.099, 0.140]
b_i^p				0.570 [0.471, 0.673]	0.018 [−0.007, 0.046]
b_i^q					0.126 [0.111, 0.142]

results show that different customers behave differently in terms of their visit patterns. For example, the mean number of visits per cluster ranges from 1.38 to 4.41. The within-cluster intervisit time ranges from 0.41 days to 1.45 days, while the between-cluster intervisit time ranges from a few weeks to more than two months. Looking at the within-cluster intervisit time and the number of visits per cluster, we find that a visit cluster spans less than a day for some customers and several days for others.

As discussed earlier, online customers' multivisit tendency for a purchase decision indicates that a series of clustered visits could serve as a reasonable proxy for a shopping episode. This in turn suggests that the conversion rate per visit cluster may serve as a useful metric to measure the efficiency of the visit-to-purchase conversion process. We calculate individuals' conversion rates per cluster and report them together with the conversion rates per visit in Table 8. We see that there is substantial variation in the conversion rates per cluster across customers, even among those who have the same conversion rate per visit. Accordingly, the customers will be scored differently depending on which conversion metric is used to evaluate their value.

Using these results on individual-level visit patterns, the online retailer could consider different marketing approaches to different customers. As shown in Table 8, some customers tend to make multiple store visits once they are in an active state of shopping, which usually lasts for a few days, and their clustered visits are more likely to result in a purchase than others. When high-clustering customers do not end up making a purchase in their first few visits within clusters, the online retailer could benefit by communicating with the customers, with customized offers and services based on their browsing activities (e.g., products searched or viewed) at those prior visits, as they are likely to come back to the retailer shortly and make additional visits within clusters. In particular, the online retailer may want to pay more attention to such customers if they have positive coefficients for S_{ij} in the purchase model (i.e., θ_{1i}^u), as it indicates that their purchase propensity tends to increase toward the end of clusters. On the other hand, when low-clustering customers do not make a purchase in their first or second visit within clusters, they are less likely to make additional visits shortly. Rather, it could take several weeks until they come back to the store and start a new visit cluster. Accordingly, for these customers, the online retailer could be better off designing marketing communications that can prompt them to initiate a new shopping episode, rather than the "follow-up" approach adopted for high-clustering customers.

Table 8 Customer-Specific Statistics of Latent Visit Clusters

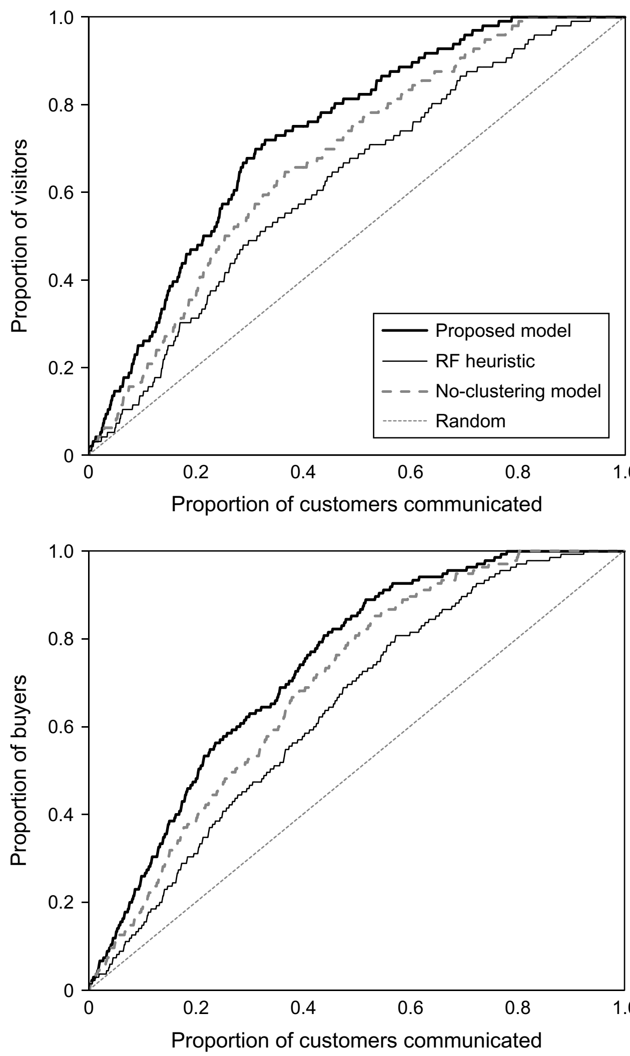
Quantile (%)	No. of visits	No. of visits per cluster	Intervisit times within a cluster	Time length of a cluster	Intervisit times between clusters	No. of clusters	Conversion rate per visit	Conversion rate per cluster
5	34	3.07	0.41	0.86	44.69	11.09	0.18	0.36
10	26	2.16	1.20	1.40	34.36	12.04	0.08	0.17
15	22	2.71	1.45	2.52	69.03	8.11	0.14	0.37
20	19	1.59	0.56	0.32	80.69	11.93	0.05	0.08
25	17	3.41	0.68	1.63	22.27	4.98	0.06	0.20
30	15	2.13	1.35	1.54	24.52	7.04	0.07	0.14
35	13	4.41	0.85	2.82	36.57	2.95	0.15	0.33
40	12	2.35	0.56	0.78	21.00	5.10	0.08	0.20
45	11	1.38	0.81	0.31	69.76	8.00	0.18	0.25
Median	10	3.13	1.44	3.37	35.37	3.20	0.10	0.33

Taking into account individuals' visit patterns and their association with purchase behavior, our model enables us to better identify future visitors and buyers. This ability of the model could help in scoring customers and making targeting decisions across customers for marketing activity. For illustrative purposes, we consider customers' visit behavior within a week after the online retailer's first email communication in the forecasting period. Using their past shopping history, we simulate individuals' probability of making a visit within a week after they receive the marketing message. We then sort the customers in decreasing order of probabilities and draw a gains chart, shown in the top panel of Figure 9. The bottom panel of Figure 9 shows another gains chart based on individuals' probability of making a purchase on their first visit after receiving the marketing communication. In both charts, the 45° line is included as a point of reference and plots the expected proportion of customers when they are selected by chance. We also consider a heuristic that ranks customers first in terms of their recency and then by frequency (RF heuristic).[17] As another point of reference, we include the predictions by the no-clustering model that does not account for clustered visit patterns and their impact on purchase behavior (i.e., Model 2 in Table 2).

In both charts, we see that the curve by our proposed model lies far above the 45° line, indicating that our model better identifies prospective customers with marketing activity undertaken by the firm. For example, in the second chart, targeting the top 50% of customers with higher purchase probabilities allows us to identify 84% of future buyers, which represents an improvement of 68% (= (0.84 − 0.50)/0.50), compared to the case of randomly selecting customers. We find that the curves by the RF heuristic also lie above the 45° line, suggesting that recency and frequency are useful predictors of customer behavior (e.g., Van Diepen et al. 2009). We also find that the no-clustering model outperforms the RF heuristic by explicitly accounting for the purchase propensity, visit rate, and latent attrition at the individual level as well as the effects of observed covariates (e.g., Schweidel and Knox 2013). Compared to the no-clustering model, our model more quickly identifies prospective customers, by utilizing their store visit patterns. These forecasts of our model can therefore assist managers in scoring customers for an efficient use of marketing resources across individuals (e.g., Venkatesan and Kumar 2004). As targeting decisions involve the computation of customers' future visit and purchase probabilities with and without marketing intervention considered at the individual level (e.g., Hansotia and Rukstales 2002), our model can help managers improve targeting decisions across individuals for marketing activity. In addition to better scoring and targeting of customers, the improved predictions of visit timing and purchase likelihood afforded by our model could assist managers' customization efforts. In particular, given technical difficulties and large costs for a complete real-time customization (e.g., Telang et al. 2004), the improved forecasts can be of great use to online retailers in personalizing website content and services beforehand, a promising way of increasing purchase conversions (e.g., Hauser et al. 2009).

[17] We also considered a heuristic in which customers were ranked first by frequency and then by recency. This alternative heuristic did not perform better than the RF heuristic we present in Figure 9.

Figure 9 **Comparison of Cumulative Gains by Alternative Methods**



6. Conclusions and Future Research

We develop an integrated model that predicts online customers' visit and purchase dynamics in a noncontractual setting. To infer store visit patterns, we assume that a customer's visit timing process possibly consists of multiple visit clusters, alternating with larger visit rates within clusters and smaller visit rates between clusters, and the propensity to remain within a cluster changes dynamically as the customer makes additional visits within the cluster. To deal with the latency of visit patterns, we employ a changepoint modeling framework and statistically infer visit patterns through data augmentation using a Bayesian approach. We then explore the question of whether (and how) predictions of online purchase and visit behaviors could be improved by accounting for visit patterns to the e-commerce site.

Using customer-level Internet clickstream data from an online retailer, we find that online store visit patterns tend to be clustered, with significant variation across customers in terms of the number and size of visit clusters and the visit frequencies, both within and between clusters. Our results show that the conversion rates vary substantially, depending on the size of a visit cluster and the location of a visit event in the cluster. Overall, the conversion rates tend to be higher at later visits within a cluster, compared with earlier visits. We also find that, after controlling for other factors that may affect purchase behavior, the visit patterns still play a significant role in predicting customers' propensity to make a purchase at a given visit. The proposed model offers superior fit and predictive performance, compared with benchmark models that ignore clustered visit patterns and their impact on purchase behavior. We demonstrate the ability of the proposed model to provide customer-specific inferences about store visit patterns and identify likely visitors and buyers, which can assist the online retailer in tailoring marketing efforts for individual customers.

Several limitations should be acknowledged and perhaps addressed in future research. First, given the focus of our research and the limitations of the data, we did not consider how customers' page-by-page browsing behavior at a given visit might influence their purchase decisions. Prior research shows that the sequence of pages viewed and the content of the pages can help improve the prediction of purchase behavior (e.g., Montgomery et al. 2004, Hauser et al. 2009). Future work might consider both browsing activities at a given visit and shopping patterns across visits in an integrated framework to better understand online shopping behavior. Second, researchers could extend our modeling framework by allowing for other latent sources of nonstationarity, in addition to that arising from clustered store visit patterns. For example, Fader et al. (2004) and Schweidel and Fader (2009) report empirical evidence of an evolving process of customer consumption patterns off-line. It would be fruitful to integrate such long-term dynamics of shopping behavior into our proposed model and disentangle their effects on observed behavior.

Third, this research has focused on studying customer shopping behavior online. There are, however, many retailers who operate both online and off-line stores. For shopping at these retailers, some customers may make use of both channels on a single purchase occasion. Restricting the analysis to an online store can therefore leave out other important touchpoints that lead to purchase conversion. Neslin et al. (2006) and Ansari et al. (2008) also suggest the importance of multichannel customer management. Another area for future research is therefore to extend our model to a multichannel setting and investigate the role of different channels in customers' shopping behavior. In a similar vein, future research may also extend our work to a multicategory context (e.g., Manchanda et al. 1999, Park et al. 2014), given that many online retailers now offer multiple product categories within a single website (e.g., Amazon.com) and customers' shopping patterns may vary across different categories. We hope that this study generates further interest and accelerates the progress in this important area of research.

Supplemental Material
Supplemental material to this paper is available at https://doi.org/10.1287/mksc.2016.0990.

Acknowledgments
The authors thank the company, which wishes to remain anonymous, that provided the data used in this study; Pete Fader, Daniel McCarthy, David Schweidel, and Daniel Zantedeschi for their valuable comments; and the reviewers for their constructive guidance and insightful suggestions. This paper stems from the first author's dissertation research, and he extends special thanks to his dissertation committee (Vishal Gaur, Sachin Gupta, and Vithala Rao) for their helpful feedback and encouragement.

References
Abe M (2009) Counting your customers one by one: A hierarchical Bayes extension to the Pareto/NBD model. *Marketing Sci.* 28(3):541–553.

Ansari A, Mela CF, Neslin SA (2008) Customer channel migration. *J. Marketing Res.* 45(1):60–76.

Bakos JY (1997) Reducing buyer search costs: Implications for electronic marketplaces. *Management Sci.* 43(12):1676–1692.

Barry D, Hartigan JA (1993) A Bayesian analysis for changepoint problems. *J. Amer. Statist. Assoc.* 88(421):309–319.

Bernasco W, Nieuwbeerta P (2005) How do residential burglars select target areas? A new approach to the analysis of criminal location choice. *British J. Criminology* 45(3):296–315.

Bucklin RE, Sismeiro C (2003) A model of Web site browsing behavior estimated on clickstream data. *J. Marketing Res.* 41(3): 249–267.

Cliff AD, Ord JK (1981) *Spatial Processes: Models and Applications* (Pion Ltd., London).

Danaher PJ (2007) Modeling page views across multiple websites with an application to Internet reach and frequency prediction. *Marketing Sci.* 26(3):422–437.

Ehrenberg ASC (1988) *Repeat-Buying: Facts, Theory and Applications* (Oxford University Press, New York).

Fader PS, Hardie BGS (2009) Probability models for customer-base analysis. *J. Interactive Marketing* 23(1):61–69.

Fader PS, Hardie BGS, Huang CY (2004) A dynamic changepoint model for new product sales forecasting. *Marketing Sci.* 23(3):280–303.

Fader PS, Hardie BGS, Lee KL (2005) Counting your customers the easy way: An alternative to the Pareto/NBD model. *Marketing Sci.* 24(2):275–284.

Forrester Research (2014) The state of retailing online 2014: Key metrics and initiatives. Forrester Research, Cambridge, MA.

Forrester Research (2015) Forrester research eCommerce forecast, 2014 to 2019 (US). Forrester Research, Cambridge, MA.

Gelman A, Rubin DB (1995) Avoiding model selection in Bayesian social research. *Sociol. Methodology* 25:165–173.

Gilovich T, Vallone R, Tversky A (1985) The hot hand in basketball: On the misperception of random sequences. *Cognitive Psych.* 17(3):295–314.

Gupta S (1988) Impact of sales promotions on when, what, and how much to buy. *J. Marketing Res.* 25(4):342–355.

Gupta S (1991) Stochastic models of interpurchase time with time-dependent covariates. *J. Marketing Res.* 28(1):1–15.

Hansotia B, Rukstales B (2002) Incremental value modeling. *J. Interactive Marketing* 16(3):35–46.

Hauser JR, Urban GL, Liberali G, Braun M (2009) Website morphing. *Marketing Sci.* 28(2):202–223.

Jain DC, Vilcassim NJ (1991) Investigating household purchase timing decisions: A conditional hazard function approach. *Marketing Sci.* 10(1):1–23.

Johnson EJ, Moe WW, Fader PS, Bellman S, Lohse J (2004) On the depth and dynamics of world wide Web shopping behavior. *Management Sci.* 50(3):299–308.

Manchanda P, Ansari A, Gupta S (1999) The "shopping basket": A model for multicategory purchase incidence decisions. *Marketing Sci.* 18(2):95–114.

Manchanda P, Rossi PE, Chintagunta PK (2004) Response modeling with nonrandom marketing-mix variables. *J. Marketing Res.* 41(4):467–478.

Manchanda P, Dubé JP, Goh KY, Chintagunta PK (2006) The effect of banner advertising on Internet purchasing. *J. Marketing Res.* 43(1):98–108.

Miller HJ, Han J (2009) *Geographic Data Mining and Knowledge Discovery* (CRC Press, Boca Raton, FL).

Moe WW (2003) Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream. *J. Consumer Psych.* 13(1–2):29–39.

Moe WW, Fader PS (2004) Dynamic conversion behavior at e-commerce sites. *Management Sci.* 50(3):326–335.

Montgomery AL, Li S, Srinivasan K, Liechty JC (2004) Modeling online browsing and path analysis using clickstream data. *Marketing Sci.* 23(4):579–595.

Moorthy S, Ratchford BT, Talukdar D (1997) Consumer information search revisited: Theory and empirical analysis. *J. Consumer Res.* 23(4):263–277.

Murray AT (2000) Spatial characteristics and comparisons of interaction and median clustering models. *Geographical Anal.* 32(1):1–18.

Neslin SA, Grewal D, Leghorn R, Shankar V, Teerling ML, Thomas JS, Verhoef PC (2006) Challenges and opportunities in multichannel customer management. *J. Service Res.* 9(2):95–112.

Netzer O, Lattin JM, Srinivasan V (2008) A hidden Markov model of customer relationship dynamics. *Marketing Sci.* 27(2):185–204.

Newton MA, Raftery AE (1994) Approximate Bayesian inference by the weighted likelihood bootstrap. *J. Roy. Statist. Soc., Ser. B* 56(1):3–48.

Park CH, Park Y-H, Schweidel DA (2014) A multi-category customer base analysis. *Internat. J. Res. Marketing* 31(3):266–279.

Park Y-H, Fader PS (2004) Browsing behavior at multiple websites. *Marketing Sci.* 23(3):280–303.

Porter JR, Howell FM (2012) *Geographical Sociology: Theoretical Foundations and Methodological Applications in the Sociology of Location* (Springer, New York).

Schmittlein DC, Bemmaor AC, Morrison DG (1985) Why does the NBD model work? Robustness in representing product purchases, brand purchases and imperfectly recorded purchases. *Marketing Sci.* 4(3):255–266.

Schmittlein DC, Morrison DG, Colombo R (1987) Counting your customers: Who are they and what will they do next? *Management Sci.* 33(1):1–24.

Schwartz EM, Fader PS, Bradlow ET (2014) Model selection using database characteristics: Developing a classification tree for longitudinal incidence data. *Marketing Sci.* 33(2):188–205.

Schweidel DA, Fader PS (2009) Dynamic changepoints revisited: An evolving process model of new product sales. *Internat. J. Res. Marketing* 26(2):119–124.

Schweidel DA, Knox G (2013) Incorporating direct marketing activity into latent attrition models. *Marketing Sci.* 32(3):471–481.

Schweidel DA, Bradlow ET, Fader PS (2009) Portfolio dynamics for customers of a multiservice provider. *Management Sci.* 57(3):471–486.

Schweidel DA, Fader PS, Bradlow ET (2008) A bivariate timing model of customer acquisition and retention. *Marketing Sci.* 27(5):829–843.

Schweidel DA, Park Y-H, Jamal Z (2014) A multiactivity latent attrition model for customer base analysis. *Marketing Sci.* 33(2):273–286.

Seetharaman PB, Chintagunta PK (2003) The proportional hazard model for purchase timing: A comparison of alternative specifications. *J. Bus. Econom. Statist.* 21(3):368–382.

Sismeiro C, Bucklin RE (2004) Modeling purchase behavior at an e-commerce Web site: A task completion approach. *J. Marketing Res.* 41(3):306–323.

Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A (2002) Bayesian measures of model complexity and fit. *J. Roy. Statist. Soc., Ser. B* 64(4):583–639.

Tanner MA, Wong WH (1987) The calculation of posterior distributions by data augmentation. *J. Amer. Statist. Assoc.* 82(398):528–540.

Telang R, Boatwright P, Mukhopadhyay T (2004) A mixture model for Internet search-engine visits. *J. Marketing Res.* 41(2):206–214.

Upton GJ, Fingleton B (1985) *Spatial Data Analysis by Example: Point Pattern and Quantitative Data* (Wiley, New York).

Urbany JE, Dickson PR, Wilkie WL (1989) Buyer uncertainty and information search. *J. Consumer Res.* 16(2):208–215.

Van Diepen M, Donkers B, Franses PH (2009) Dynamic and competitive effects of direct mailings: A charitable giving application. *J. Marketing Res.* 46(1):120–133.

Venkatesan R, Kumar V (2004) Customer lifetime value framework for customer selection and optimal resource allocation strategy. *J. Marketing* 68(4):106–125.

Zhang Y, Bradlow ET, Small DS (2013) New measures of clumpiness for incidence data. *J. Appl. Statist.* 40(11):2533–2548.

Zhang Y, Bradlow ET, Small DS (2015) Predicting customer value using clumpiness: From RFM to RFMC. *Marketing Sci.* 34(2):195–208.