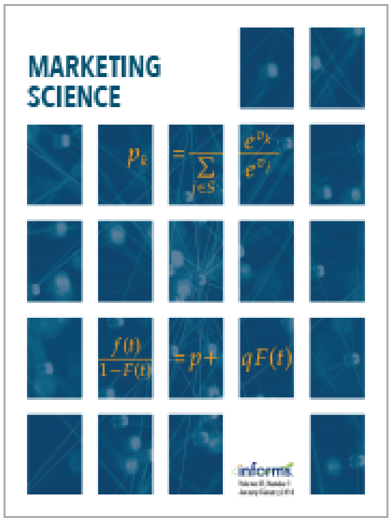


Marketing Science

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

A Multiactivity Latent Attrition Model for Customer Base Analysis

David A. Schweidel, Young-Hoon Park, Zainab Jamal

MARKETING SCIENCE
Vol. 33, No. 2, March–April 2014, pp. 273–286
ISSN 0732-2399 (print) | ISSN 1526-548X (online)

http://dx.doi.org/10.1287/mksc.2013.0832

A Multiactivity Latent Attrition Model for Customer Base Analysis

David A. Schweidel
Goizueta Business School, Emory University, Atlanta, Georgia 30322, dschweidel@emory.edu

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

Zainab Jamal
Hewlett-Packard Labs, Palo Alto, California 94304, zainab.jamal@hp.com

Customer base analysis is a key element in customer valuation and can provide guidance for decisions such as resource allocation. Yet extant models often focus on a single activity, such as purchases from a retailer or donations to a nonprofit organization. These models do not consider other ways that an individual may engage with an organization, such as purchasing in multiple brands or contributing user-generated content. In this research, we propose a framework to generalize extant models for customer base analysis to multiple activities.

Using the data from a website that allows users to purchase digital content and/or post digital content at no charge, we develop a flexible "buy 'til you die" model to empirically examine how the two activities are related. Compared with benchmarks, our model more accurately forecasts the future behavior for both types of activities. In addition to finding evidence of coincidence between the activities while customers are "alive," we find that the latent attrition processes are related. This suggests that conducting one type of activity is informative of whether customers are still alive to conduct another type of activity and, consequently, affects inferences of customer value.

Keywords: customer base analysis; latent changepoint model; multivariate choice model; multivariate "buy 'til you die" model; digital content

History: Received: April 12, 2011; accepted: November 5, 2013; Preyas Desai served as the editor-in-chief and Scott Neslin served as associate editor for this article. Published online in *Articles in Advance* January 31, 2014.

1. Introduction

Research in customer base analysis serves as the foundation for customer valuation (e.g., Schmittlein et al. 1987, Reinartz and Kumar 2000). Such models allow marketers to make forecasts at both the individual customer level and the aggregate level (e.g., Fader and Hardie 2009), and they have been applied to various aspects of industry, including retail transactions, donations, and nonprofit organizations (e.g., Fader et al. 2005, Abe 2009, Neslin et al. 2009, Singh et al. 2009). Yet one of the salient limitations of extant research in customer base analysis is that it primarily focuses on a single type of transactional activity conducted by customers.

Consider the case of a firm that owns multiple brands and operates distinct retail stores for each of its brands, such as Gap Inc. and Limited Brands. The firm can project future customer behavior for each brand individually using models such as the Pareto/NBD (Schmittlein et al. 1987) or beta-geometric/beta-Bernoulli (BG/BB) (Fader et al. 2010). Such analyses, however, fail to consider the possible associations that exist in purchasing behavior across the brands. By looking at the customer's transaction histories across different brands, the firm can assess those brands that a customer is still likely to purchase in the future. Doing so may reveal, for example, that continued purchasing of one brand is related to the continued purchasing of another brand, aiding the firm in its targeting and communication strategies.

A similar situation arises for online vendors that sell products in multiple categories, such as event tickets in categories including concerts and sports (e.g., http://www.StubHub.com). Although the vendor may separately analyze a customer's past purchasing of concert tickets to project his future purchasing of concert tickets, the customer's relationship with the company may involve purchasing both concert and sport tickets. By examining these purchase histories jointly, the company can determine whether he is prone to cease purchasing from it entirely. Discussions with a Fortune 500 firm that only operates online and sells products in multiple verticals revealed a similar concern, with management expressing a strong interest in understanding if churn in one vertical is informative of future churn in other verticals or if customers churn across all verticals at once. Although multiple data streams have been incorporated into multivariate choice models (e.g., Manchanda et al. 1999), limited research in customer base analysis has examined such cross-activity effects.

Whereas these examples describe scenarios with multiple revenue streams, there are also contexts in which multiple types of transactional activities are observed with only a single revenue stream. For example, online artist communities allow visitors to post comments about artists' work, as well as purchase prints and other products featuring the work (e.g., http://www.deviantART.com). This is also the case with donor development, where alumni may be more likely to make a contribution to their alma mater after attending a reunion or volunteering (e.g., Netzer et al. 2008). Although individuals volunteering or attending reunions do not directly generate revenue for the university, they may increase the likelihood or volume of subsequent giving activity. Online product ratings, which have been linked to increased purchasing (e.g., Moe and Trusov 2011), also provide an illustration of customers engaging in multiple types of activities (purchasing and providing ratings) that may be linked to each other.

In this research, we generalize the "buy 'til you die" models often employed in customer base analysis (e.g., Schmittlein et al. 1987, Fader et al. 2010) to a multiactivity setting. Rather than assuming that one activity is an exogenous driver of another activity, we jointly model the different activities that a customer may conduct. Consistent with the multivariate choice literature, we account for both correlation among customers' tendencies to conduct different types of transactional activities and coincidence among the activities (e.g., Manchanda et al. 1999, Park and Fader 2004). We also incorporate a multivariate latent attrition process that incorporates correlation in customers' tendencies to become inactive and links the processes themselves. The result is a multivariate buy 'til you die model that nests extant models for customer base analysis.

To the best of our knowledge, our research is among the first to explore the question of how latent attrition occurs in a multivariate context. One possibility is that customers become inactive on individual activities, one at a time. In such a scenario, the latent attrition processes are (conditionally) independent. An alternative to this is that still being active on one activity affects the tendency to become inactive on another activity. In this case, the latent attrition processes for each activity exhibit a dependency, and inferring that a customer is still active on one activity will affect expectations of future behavior of another activity. Customers may also exhibit coincidence in their attrition on different activities, with customers having an increased tendency to become inactive on all activities at once. In the extreme, coincidence in the latent attrition processes would manifest as customers become inactive with respect to all transactional activities at once rather than progressing through a staged attrition process.

To examine the nature of latent attrition in a multiactivity setting, we demonstrate the applicability of our modeling framework using data provided by an e-commerce website that allows users to purchase digital content and/or post digital content so that it may be viewed by others. Our empirical findings suggest that the attrition processes between the two activities are linked to each other, with customers who are active on both activities being less likely to become inactive on either activity compared with customers who are already inactive on a single activity. Compared with a set of benchmark models, we find that our modeling framework more accurately predicts customers' future behavior for the majority of customers.

Although extant buy 'til you die models have provided insights for scoring customers and predicting customers' future transactions with regard to a single type of transactional activity, they have fallen short of providing such guidance in the presence of multiple types of transactional activities, such as purchasing from multiple brands or categories. We show how predictions of the activities in which customers may engage in the future are affected by multiple activities, highlighting the potential for improved targeting based not just on the recency and frequency of a single activity but also on the set of activities a customer has previously conducted. In doing so, managers may leverage customer-level data from multiple activities to address the question of when they should "give up" on customers who have not conducted any activities in some time. As we will show, inferences of customers' future behavior for those who have conducted no transactions differ dramatically from inferences about customers who have engaged in some activities. Such environments have become more prevalent as online retailers encourage customers to engage in social media activity relating to their purchases. As the overlap between customers' social media activities and firm's customer relationship management efforts continues to grow, our modeling framework offers a means by which firms can combine customers' past purchasing and social media-contributing behavior to predict their future purchasing.

The remainder of this paper proceeds as follows. We next describe the data used in our empirical application. In §3, we develop a dynamic bivariate latent attrition model. Throughout, we note the ease with which our modeling framework generalizes beyond two activities. In §4, we discuss the model results and illustrate how our framework can be used by firms in managing relationships with customers in a multiactivity setting. Section 5 concludes with a discussion of our key findings and proposes directions for future research.

2. Data

Data for our empirical application were provided by an e-commerce website that allows customers to engage in two different types of transactional activities. The first activity in which customers may engage is purchasing digital content. The second activity in which customers may engage is the posting of digital content for viewing by others. This activity is offered to customers at no cost and is intended to increase customers' engagement with the website. In contrast to sequential activities, which require the completion of one activity for a customer to engage in the other activity (e.g., Sismeiro and Bucklin 2004), purchasing and posting are not sequential by design. Because the activities are not sequential by design, subsequent to registering at the e-commerce website, customers may make multiple purchases without ever posting content or may exclusively use the posting activity. A random sample of 5,000 customers was selected from the early 2000s and tracked for 52 weeks after their registration. Our data indicate each week whether or not a customer had a transaction involving each activity. We used the first 26 weeks to calibrate the model and the remaining 26 weeks for model validation, providing a rigorous test of the modeling framework's forecasting ability.[1]

Consistent with the industry in which the data provider operates, a substantial fraction of customers register with the website and are not observed to conduct transactions of either type during our observation period (83%). Because the activities are not sequential in nature, customers may engage in only one activity, both activities, or neither activity. Among the customers who are observed to conduct transactions, 61% engage in both activities, whereas 28% only purchase and 11% only post. During the 52-week data period, customers who both purchased and posted had an average of 4.6 weeks in which purchases occurred and 3.6 weeks in which they used the posting activity. In contrast, those customers who only made purchases did so in 2.3 (on average) of the 52 weeks, and those who only used the posting activity did so in 1.7 weeks. Although this initial analysis suggests that users engaged in both activities are heavier users than individuals who only engage in a single activity, potentially as a result of the presence of correlated propensities, it offers limited insight into the types and magnitudes of association between the two activities.

Table 1 **Distribution of Activities Conditional on Observing Any Activity**

Weeks with purchases	Weeks with postings (%)			
	0	1	2–5	>5
0	—	8.0	3.2	0.2
1	16.4	8.4	4.0	0.2
2–5	9.6	8.0	18.7	2.3
>5	2.0	2.2	8.0	8.7

To explore customers' joint behavior initially, we consider the frequency with which customers engage in the two activities. In Table 1, conditional on engaging in at least one activity during the data period, we classify customers based on the number of weeks in which they conducted each activity.[2] We see a clear pattern emerge in Table 1, suggesting an association between the frequency of purchasing and posting. Of the customers observed to conduct at least one transaction, a large group of customers use both activities no more than once (32.8%). Those customers who have a moderate purchase incidence (i.e., two to five weeks with purchases) also tend to have a moderate posting incidence, and customers who frequently purchase (i.e., more than five weeks with purchases) also use the posting activity frequently.

Although Table 1 indicates the presence of a relationship between the purchasing and posting activities, it does not speak to the coincidence in the timing of the activities. Moreover, it does not allow us to explore the relationship between the latent attrition processes. To examine these aspects of customer behavior in a multiactivity context, we look at the time at which purchases are made compared with the time the nearest posting was made for those customers who are observed to engage in both activities. Of all purchases made by these customers, 31.7% are made the same week in which they posted content, and 65.1% are made within two weeks of using the posting activity. Table 2 also suggests that customers exhibit coincidence in their transactions, with 70.5% purchasing the same week they post content and 95.5% making a purchase within two weeks of making a post. This pattern may also suggest a relationship between the latent attrition processes, as customers active on (and conducting)

[1] In the industry in which this e-commerce site operates, transactions occur relatively infrequently, making the weekly level a reasonable unit of analysis. An initial analysis was conducted with the daily data, which yielded qualitatively similar results to those reported in this paper. Based on conversations with the data provider, we decided to conduct the analysis with the weekly data in order to correspond to the way in which the firm monitors key metrics regarding customer activities.

[2] Because of the sparseness of the transactions in the two activities considered in this research, we have aggregated the number of weeks to illustrate the potential association in the joint activities. Our empirical analysis and results assume no such aggregation.

Table 2 Coincidence Between Purchasing and Posting

Nearest use of posting	Percentage of purchases	Percentage of customers
Same week	31.7	70.5
Within one week	56.8	92.2
Within two weeks	65.1	95.5

posting are likely still active with respect to purchasing, necessitating a joint analysis of the two activities.

3. Model Development

We build a dynamic bivariate model of choice behavior that accounts for latent attrition in customers' use of each type of transactional activity. We begin by developing a dynamic bivariate choice model for customers' decisions of which activities to conduct each week, conditional on customers being active on each activity, in §3.1. We then allow for latent attrition with respect to customers' use of each activity in §3.2. We discuss our accommodation of unobserved heterogeneity across customers in §3.3.

3.1. A Dynamic Bivariate Choice Model

We begin by developing a bivariate choice model for customers' purchasing and posting. In doing so, we account for heterogeneity, the impact of prior activity use on subsequent activity use, and coincidence. If customer h uses of activity k in week t, we denote $y_{htk}=1$; otherwise, $y_{htk}=0$. We assume that customer h's utility for using activity k in week t is given by

$$u_{htk}=\alpha_{hk}+\sum_{j=1}^{2}\beta_{kj}\cdot Stock_{htj} \quad (1)$$

for $k=1$ or 2. The customer-by-activity intercept α_{hk} captures heterogeneity in customers' tendencies to use activity k. The coefficients β_{kj} allow for the tendency to engage in activity k to vary over time and be related to the recency and frequency with which activity j was conducted, which is reflected by $Stock_{htj}$. This specification allows for $Stock_{htj}$ to exhibit a different impact on the tendency to conduct activity 1 (reflected by β_{1j}) and on the tendency to conduct activity 2 (β_{2j}). We specify $Stock_{htj}$ as

$$Stock_{htj}=\delta_j\, Stock_{h,\,t-1,\,j}+y_{h,\,t-1,\,j}, \quad (2)$$

where δ_j is an activity-specific decay factor between 0 and 1. Note that $Stock_{htj}$ may increase with recent transactions of activity j or decrease because of the decay process (δ_j). As more time elapses without any activities being conducted, $Stock_{htj}$ approaches 0 for all j, and the utility associated with conducting activity k will approach α_{hk}. By assuming that the customer-by-activity intercepts are time invariant, we can distinguish the heterogeneity across customers (captured by α_{hk}) from the temporal variation in customers' tendencies to conduct transactions related to their prior activity (captured by β_{kj} and δ_j).

This specification allows for a number of ways in which customers' behavior may fluctuate over time based on customers' recent activities. If customers are more likely to purchase in the weeks immediately after having done so, then $\beta_{11}>0$. Alternatively, if $\beta_{11}<0$, this may be akin to customers being "stocked up" from recent activity use and hence less likely to conduct activities in the immediate future (e.g., Ailawadi and Neslin 1998, Bell et al. 1998). Under this scenario, as more time elapses since the last purchase, $Stock_{ht1}$ will diminish, and the likelihood with which a customer conducts a purchase will increase. Similarly, we may investigate the cross effects, such as the effect of posting content on future purchases. For instance, if customers are more likely to make purchases in the weeks immediately following a posting activity, then $\beta_{12}>0$.

To account for the potentially increased utility associated with purchasing and posting activity in the same week, resulting in coincidence, we employ a bivariate logit model (e.g., Russell and Petersen 2000, Niraj et al. 2008). If customer h were to engage in activity k in week t and not engage in any other activity, the deterministic component of his utility is given by u_{htk}. If customer h were to both purchase and post, we assume that the deterministic component of his utility is given by $u_{th1}+u_{ht2}+\theta$, where θ reflects the change in utility associated with conducting both activities in the same week. Let $z_{htk}=1$ be a latent variable to indicate that customer h is still active on activity k at time t; let it be equal to 0 otherwise. Conditional on $z_{htk}=1$ for $k=1$ and $k=2$, the probabilities corresponding to the combination of activities in which customer h engages in week t are then given by

$$p_{ht}(I_1,I_2\mid z_{ht1}=1,z_{ht2}=1) =\frac{\exp(I_1u_{ht1}+I_2u_{ht2}+I_1I_2\theta)}{1+\exp(u_{ht1})+\exp(u_{ht2})+\exp(u_{ht1}+u_{ht2}+\theta)}, \quad (3)$$

where $I_k=1$ if customer h uses activity k in week t and 0 otherwise.[3] Based on the different combinations of values for I_1 and I_2, Equation (3) yields four distinct

[3] Although our empirical application focuses on the bivariate context, prior research has applied the multivariate logistic choice model (Cox 1972) to the n-variate context for $n>2$. This is achieved by generalizing the denominator of Equation (3) to sum over the 2^N terms corresponding to the deterministic component of the utility for each possible combination of $(I_1,I_2,\ldots,I_N)$. For a marketing example of the multivariate logistic model, we refer interested readers to Russell and Petersen (2000), who employ a multivariate category choice model with $n=4$ and introduce pairwise dependencies across categories through additional θ terms.

probabilities for customer h at time t conditional on $z_{ht1}=1$ and $z_{ht2}=1$: $p_{ht}(1,1)$, $p_{ht}(0,0)$, $p_{ht}(1,0)$ and $p_{ht}(0,1)$.

It is through the term θ that the bivariate logit specification presented in Equation (3) captures coincidence in the activities (Manchanda et al. 1999, Park and Fader 2004). If $\theta=0$, the probability of purchasing *and* posting content in week t can be expressed as the product of the purchasing probability and the posting probability, with each probability corresponding to a binary logit model. Values of $\theta>0$ ($\theta<0$) imply that purchasing and posting content are more (less) likely to occur within the same week than independence would suggest. Although this specification does not allow us to address the potential causal link between purchasing and posting activities, we do accommodate coincidence in the activities. In-process measures such as clickstream data pertaining to customers' browsing activities leading up to transactions could perhaps allow for the causal relationship to be explored more fully.

Coincidence between purchasing and posting may exist while a customer is active on both purchasing ($z_{ht1}=1$) and posting ($z_{ht2}=1$). We next consider the case in which customer h is inactive on one activity. Suppose that $z_{ht2}=0$, and customer h is only active with respect to purchasing. The probability of customer h posting ($k=2$) at time t is therefore equal to 0, and the probability of customer h purchasing at time t is then given by the binary logit model with utility u_{ht1}. The corresponding probabilities for the vector of observations (I_1, I_2) can then be written as

$$\begin{aligned} p_{ht}(I_1,I_2\mid z_{ht1}=1,z_{ht2}=0)&=\frac{\exp(I_1u_{ht1})}{1+\exp(u_{ht1})}, \quad I_2=0,\\ p_{ht}(I_1,I_2\mid z_{ht1}=1,z_{ht2}=0)&=0, \quad I_2=1. \end{aligned} \tag{4}$$

Similarly, if $z_{ht1}=0$ and $z_{ht2}=1$, the probabilities corresponding to the vector of observations (I_1, I_2) is given by

$$\begin{aligned} p_{ht}(I_1,I_2\mid z_{ht1}=0,z_{ht2}=1)&=\frac{\exp(I_2u_{ht2})}{1+\exp(u_{ht2})}, \quad I_1=0,\\ p_{ht}(I_1,I_2\mid z_{ht1}=0,z_{ht2}=1)&=0, \quad I_1=1. \end{aligned} \tag{5}$$

If $z_{htk}=0$ for both $k=1$ and $k=2$, then the only possible observation at time t is that customer h neither purchases nor posts ($p_{ht}(0,0\mid z_{ht1}=0, z_{ht2}=0)=1$).

If data were available about the firm's marketing efforts, such efforts can easily be incorporated into the dynamic bivariate choice model through u_{htk} as described in Equation (1). Marketing efforts focused on purchasing could directly be incorporated into u_{ht1}, and the cross effects of marketing could be incorporated into u_{ht2}. Similarly, marketing focused on encouraging the posting activity could be incorporated into u_{ht2}, and the cross effects incorporated into u_{ht1}. The resulting model would encompass own and cross effects from marketing, coincidence, and heterogeneity across customers, thus capturing the same own- and cross-category dependencies as extant multivariate choice models (e.g., Manchanda et al. 1999).

The model presented in Equations (1)–(5) describes the likelihood with which customers engage in both activities, conditional on the set of activities on which they are still active. We next detail our model for the latent transition between active and inactive states for each activity.

3.2. The Latent Attrition Process

We adopt a similar model specification to capture both variations across customers and over time in the latent attrition process. We assume that customers may become inactive on each activity and transition to an absorbing "death" state with respect to that activity after each time period (e.g., Fader et al. 2010). Similar to Equation (1), we define q_{htk} as customer h's tendency to become inactive with respect to activity k at time t:

$$q_{htk}=\kappa_{hk}+\sum_{j=1}^{2}\psi_{kj}\cdot Stock_{htj}, \tag{6}$$

where the intercept κ_{hk} captures heterogeneity across customers in their tendency to become inactive users of activity k, and ψ_{kj} reflects the degree to which this tendency may change based on a customer's recent transactions of activity j. Depending on the value of ψ_{kj}, customers may become more or less likely to remain active on activity k.[4] If $\psi_{kj}=0$, customer h's tendency to become inactive on activity k would be time invariant (e.g., Fader et al. 2010). If other time-varying covariates were available and believed to affect the transaction incidence or attrition processes, they could be incorporated into Equations (1) and (6) (e.g., Schweidel and Knox 2013).

Although Equation (6) allows for heterogeneity across customers and the effect of recent activity on the attrition processes, it omits two important factors that may affect the activity-specific attrition processes in a multiactivity setting. First, there may be a coincidence in the times at which customers become inactive on different activities. That is, customers may exhibit a tendency to become inactive on multiple activities at the same time beyond that which would occur if the attrition processes were independent. To capture this form of interdependence, we incorporate the parameter ω to reflect the increased tendency for customers to become inactive on both purchasing and posting simultaneously.

[4] In the model estimated with heterogeneous parameters β_{hk}, we also allowed for heterogeneity in the extent to which prior activity affects the attrition process (ψ_{hk}). As noted previously, this model did not perform better during the forecasting period and yielded similar substantive findings to the specifications we present.

Second, the tendency for customers to become inactive on one activity may be linked to other activities, depending on whether they have already become inactive on other activities. For example, take the case of a retailer that sells in multiple brands. If a customer has ceased purchasing one brand, this may be indicative of a weakening relationship with the retailer that contributes to an increased tendency to become inactive on other brands. This would be consistent with prior research that has found that the number of categories in which customers purchase is associated with reduced customer attrition (e.g., Reinartz and Kumar 2003). To allow for the possibility that being active on other activities affects the latent attrition process, we incorporate the parameter τ_1 to reflect the difference in the tendency to become inactive on purchasing while a customer is still active on posting and τ_2 to reflect the difference in the tendency to become inactive on posting while a customer is still active on purchasing.

If a customer is still active on both purchasing ($z_{ht1}=1$) and posting ($z_{ht2}=1$), we let $r_{ht}(I_1, I_2)$ define the probability of becoming inactive on purchasing ($I_1=1$ and 0 otherwise) and/or posting ($I_2=1$ and 0 otherwise) as

$$r_{ht}(I_1, I_2 \mid z_{ht1}=1, z_{ht2}=1) = \frac{\exp(\tilde{r}_{ht}(I_1, I_2))}{\exp(\tilde{r}_{ht}(1,1))+\exp(\tilde{r}_{ht}(1,0))+\exp(\tilde{r}_{ht}(0,1))+\exp(\tilde{r}_{ht}(0,0))}, \quad (7)$$

where

$$\begin{aligned} \tilde{r}_{ht}(1,1) &= (q_{ht1}+\tau_1)+(q_{ht2}+\tau_2)+\omega, \\ \tilde{r}_{ht}(1,0) &= (q_{ht1}+\tau_1), \\ \tilde{r}_{ht}(0,1) &= (q_{ht2}+\tau_2), \\ \tilde{r}_{ht}(0,0) &= 0. \end{aligned} \quad (8)$$

When a customer is only active on an individual activity, there is no consideration of coincidence in the attrition processes. If customer h is still active on purchasing ($z_{ht1}=1$) but not posting ($z_{ht2}=0$), the probability of becoming inactive on purchasing is given by $\text{logit}^{-1}(q_{ht1})$. Similarly, if customer h is still active on posting but not purchasing, the probability of becoming inactive on posting is given by $\text{logit}^{-1}(q_{ht2})$.

Substituting Equation (6) into Equation (8), the tendency to become inactive with respect to activity k is given by $\kappa_{hk}+\tau_k+\sum \psi_{kj} \cdot Stock_{htj}$. We can identify the coefficients ψ_{kj} based on variation in the stock variables that occurs over time. To distinguish κ_{hk} and τ_k, recall that τ_k only plays a role in the attrition process when customers are active on both activities. In contrast, the customer-specific intercept κ_{hk} affects the attrition process when customers are active on both activities, as well as when customers are only active on a single activity. An alternative way to construct the transition probabilities would be to specify one intercept for the tendency to become inactive on activity k when customers are only active on activity k and a separate intercept for the tendency to become inactive on activity k when customers are still active on multiple activities. Rather than specify two distinct intercepts, we explicate our model by specifying a customer-specific intercept (κ_{hk}) and let τ_k reflect the difference in the tendency to become inactive on activity k when active on multiple activities. It should also be noted that τ_k and ψ_{kj} capture different aspects of customers' tendency toward attrition on activity k. Whereas the impact of τ_k is time invariant and dependent on whether customers are still active on both activities, the impact of ψ_{jk} varies with the current level of $Stock_{htj}$, which depends on the recency and frequency with which customer h has conducted activity j at time t.

The specification for the latent attrition process described previously captures a number of important features, many of which have not previously been documented in the customer base analysis literature. In addition to allowing for the attrition process to vary across customers and over time, we allow for the processes to be linked to each other in two ways. We also allow for customers' tendencies to become inactive to depend on whether they are still active on other activities, captured by the parameter vector τ. Such information is essential to multiactivity firms, as the lack of transactions of one activity can serve as a harbinger of a customer who is likely to become inactive on other activities. Moreover, we incorporate coincidence in latent attrition. Should coincidence in attrition be present among activities, firms would be at risk of losing customers on multiple activities simultaneously.

Our specification for the latent attrition processes nests simpler models that assume the attrition processes are independent. If $\omega > 0$, this would suggest that customers are likely to cease both activities at once, consistent with an "all or nothing" attrition process (Schweidel et al. 2011). If we assume that $\omega = 0$, then Equation (7) can be expressed as the product of two binary logit probabilities where the likelihood of becoming inactive on activity k is given by $\text{logit}^{-1}(q_{htk}+\tau_k)$. If we were to further assume that $\tau_1=\tau_2=0$, the latent attrition probabilities for each activity would not depend on customers being active on the remaining activities. If being active on multiple activities is a proxy for customer engagement, we would expect $\tau_k < 0$, suggesting that attrition is less likely on activity k when customers are active on other activities.

Our intent in the current research is to contribute to our understanding of how latent attrition occurs in a multiactivity setting. Whereas we investigate a bivariate setting, the ideas of linked attrition (τ_1 and τ_2) and coincidence (ω) are relevant beyond the bivariate context. To maintain parsimony while generalizing the latent attrition process to an n-variate setting, one could to specify τ using a parametric function of the number of other activities on which a customer is still active. If one views coincidence in the latent attrition process as suggesting that all activities are more inclined to cease at once, no additional model parameters beyond ω are necessary. A more flexible view would assume that certain activities may experience attrition together. To account for this, one could include pairwise and trivariate association terms; for example, the pairwise terms ω_{ab}, ω_{ac}, ω_{bc} and trivariate term ω_{abc} could be used to approximate the associations among the latent attrition processes for activities a, b, and c (Danaher 2007).

3.3. Incorporating Heterogeneity Across Customers

To complete our specification, we next describe how we account for heterogeneity across customers. First, we assume that customers may begin as active on any combination of activities, with the initial distribution given by the vector π. In our empirical application, this yields four possibilities: customers are active on (1) purchasing and posting, (2) only purchasing, (3) only posting, or (4) none of the activities. We let the nth element of the vector π indicate the probability with which customers begin active on the aforementioned sets of activities subsequent to registration at the website. Conditional on not starting in the inactive state, we allow for differences across customers with respect to their tendencies to purchase and post content, as well as the tendency to remain active on each activity. To do so, we assume that the customer-level intercepts in Equations (1) and (6) follow a multivariate normal distribution:

$$\begin{pmatrix} \alpha_{h1} \\ \alpha_{h2} \\ \kappa_{h1} \\ \kappa_{h2} \end{pmatrix} \sim \mathrm{MVN}(\Phi, \Sigma). \tag{9}$$

This specification allows for the underlying tendencies to conduct the different activities, as well as the tendency to remain in an active relationship, to be correlated (e.g., Abe 2009).

As the model described in Equations (1)–(9) is a latent changepoint model similar in spirit to a hidden Markov models (HMMs), we must consider the different possible times at which individuals may transition from active to inactive on each activity. There are four possible states that we consider, based on being active with respect to purchasing and posting. In our empirical application, we let the state representing the combination of activities on which customer h is still active be given by $s = 1 + 2(1 - z_{ht1}) + (1 - z_{ht2})$. As such, a customer in the first state is active on both purchasing and posting ($z_{ht1} = z_{ht2} = 1$), a customer active only on purchasing ($z_{ht1} = 1$ and $z_{ht2} = 0$) is in the second state, a customer active only on posting ($z_{ht1} = 0$ and $z_{ht2} = 1$) is in the third state, and a customer inactive on both activities ($z_{ht1} = z_{ht2} = 0$) is in the fourth state. The likelihood function for customer h can be expressed as follows (MacDonald and Zucchini 1997):

$$L_h = \boldsymbol{\pi} D_{h1} W_{h1} D_{h2} W_{h2} \cdots W_{h,T-1} D_{hT} \mathbf{1}, \tag{10}$$

where $\boldsymbol{\pi}$ is the initial distribution, W_{ht} is the transition matrix, $\mathbf{1}$ is a 4×1 vector of ones, and D_{ht} is a diagonal matrix composed of the likelihoods conditional on the activities on which customer h is active:

$$D_{ht} = \begin{pmatrix} a_{11} & 0 & 0 & 0 \\ 0 & a_{22} & 0 & 0 \\ 0 & 0 & a_{33} & 0 \\ 0 & 0 & 0 & a_{44} \end{pmatrix}, \tag{11}$$

where

$$\begin{aligned} a_{11} &= p_{ht}(y_{ht1}, y_{ht2} \mid z_{ht1} = 1, z_{ht2} = 1), \\ a_{22} &= p_{ht}(y_{ht1}, y_{ht2} \mid z_{ht1} = 1, z_{ht2} = 0), \\ a_{33} &= p_{ht}(y_{ht1}, y_{ht2} \mid z_{ht1} = 0, z_{ht2} = 1), \\ a_{44} &= (1 - y_{ht1})(1 - y_{ht2}). \end{aligned}$$

The first diagonal element of D_{ht} is the likelihood conditional on being active on both activities (Equation (3)), the second element is the likelihood conditional on being active on purchasing only (Equation (4)), the third element is the conditional likelihood associated with only being active on posting (Equation (5)), and the final diagonal element is the likelihood conditional on being inactive on both activities.

The transition matrix W_{ht} accounts for the different activities on which a customer is still active. The first row includes the transition probabilities for a customer who is active on both activities, the second row has the probabilities for customers only active on purchasing, the third row contains the probabilities for customers only active on posting, and the final row assumes that the inactive state is absorbing:

$$W_{ht} = \begin{pmatrix} w_{11} & w_{12} & w_{13} & w_{14} \\ 0 & w_{22} & 0 & w_{24} \\ 0 & 0 & w_{33} & w_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{12}$$

where

$$\begin{aligned} w_{11} &= r_{ht}(0, 0 \mid z_{ht1} = 1, z_{ht2} = 1), \\ w_{12} &= r_{ht}(0, 1 \mid z_{ht1} = 1, z_{ht2} = 1), \\ w_{13} &= r_{ht}(1, 0 \mid z_{ht1} = 1, z_{ht2} = 1), \\ w_{14} &= r_{ht}(1, 1 \mid z_{ht1} = 1, z_{ht2} = 1), \\ w_{22} &= 1 - \mathrm{logit}^{-1}(q_{ht1}), \quad w_{24} = \mathrm{logit}^{-1}(q_{ht1}), \\ w_{33} &= 1 - \mathrm{logit}^{-1}(q_{ht2}), \quad w_{34} = \mathrm{logit}^{-1}(q_{ht2}). \end{aligned}$$

When active on both purchasing and posting, a customer may transition to any of the states by remaining active on both activities (staying in the first state), becoming inactive on posting (moving to the second state), becoming inactive on purchasing (moving to the third state), or becoming inactive on both activities (moving to the absorbing fourth state). From either the second or third states, though, the only possibilities are staying active on the remaining activity or become inactive on the activity and moving to the absorbing state. Beyond the accommodations already discussed to generalize the bivariate incidence and latent attrition models to more activity types, although the dimensions of W_{ht} increase with the number of activities, this only requires additional "bookkeeping" of terms to accommodate the possible transitions among the latent states (Park and Fader 2004).

4. An Empirical Application

We estimate the proposed model (as specified in Equations (1)–(12)) using a hierarchical Bayesian model. We assume diffuse normal priors for the fixed effects that do not vary across customers (Φ, θ, ω, τ, β, ψ, and logit(δ)). We assume a diffuse inverse Wishart distribution for Σ. We estimate the model using a Metropolis-Hastings algorithm, drawing samples from the posterior distribution of 25,000 draws from two independent Markov chain Monte Carlo (MCMC) chains following a burn-in of 25,000 iterations.[5]

In addition to the proposed model, we estimate three important benchmark models. First, we estimate two independent BG/BB models (Fader et al. 2010), one corresponding to purchasing and one corresponding to posting. Although this model accounts for customers' transitions to an inactive absorbing state for each activity, it does not consider the potential own or cross effects of recent activity use. The second benchmark model that we employ is a bivariate logit choice model with time-varying covariates and customer-by-activity–specific intercepts. We include the time since each activity was last used (recency) and the number of weeks in which each activity has been used (frequency) prior to week t as predictors, as well as the effect of coincidence. The utility associated with customer h conducting activity k in week t is given by

$$u_{htk} = a_{hk} + \sum_{j=1}^{2} \beta_j \cdot Recency_{htj} + \sum_{j=1}^{2} \gamma_j \cdot Frequency_{htj}. \quad (13)$$

Finally, we estimate a benchmark model that assumes customers exhibit a single latent attrition process. In contrast to the latent attrition processes described in Equations (6)–(8), which allow for the possibility that a customer becomes inactive on one activity but remains active on another, this benchmark model assumes that customers are either active on both activities or inactive on both activities. Comparing this model with a single attrition process to our proposed model, we assess the improved forecasting ability of incorporating activity-specific latent attrition processes.

4.1. Model Comparison

In Table 3, we report the log marginal density (LMD) of the proposed model, as well as the LMD of the bivariate logit benchmark with recency and frequency (RF) covariates and the independent BG/BB benchmark. To assess the forecasting performance of the models, we split our data into a 26-week calibration period and a 26-week forecasting period. During the latter, we simulate the future behavior for each customer in our data set week by week at each iteration of the MCMC sampler and then average across the iterations. From these simulations, we constructed tracking plots of the number of customers expected to conduct each activity each week of the holdout period. We also present the mean absolute percentage error (MAPE), averaged across weeks, in Table 3.

On the basis of LMD during the calibration period, whereas our proposed model outperforms the independent BG/BB benchmark, the bivariate logit and single attrition benchmark models outperform the proposed model. This is due to, in part, the additional complexity of the proposed model relative to these benchmarks, as the purchasing and posting MAPEs reveal relatively close performance among the alternative specifications. A different story emerges when we consider the holdout period. For both purchasing and posting behavior, we see that our proposed model outperforms the three benchmarks. The tracking plots are presented in Figure 1.

The vertical dashed line indicates the end of the calibration period. As can be seen with respect to both purchasing and posting, the model captures the general trend during both the calibration and forecasting periods. The bivariate logit model with recency and frequency variables generally underpredicts purchasing and posting transactions during the forecasting period, despite capturing these trends during the calibration period. In contrast, the single attrition benchmark model, while also reflecting the general transaction patterns during the calibration period, tends to overpredict both types of transactions. Although the independent BG/BB model closely tracks our proposed model for posting transactions, we see that it tends to overpredict purchase occasions.

Whereas Figure 1 examines the model's performance aggregated to the weekly level, we next consider the

[5] The estimation procedure follows that employed by recent research using latent changepoint models (Netzer et al. 2008, Schweidel et al. 2011). Additional details are available from the authors upon request.

Table 3 Model Comparison

		Calibration period		Holdout period	
Model	LMD	Purchasing MAPE (%)	Posting MAPE (%)	Purchasing MAPE (%)	Posting MAPE (%)
Proposed model	−11,759	9.5	10.9	12.2	17.1
Independent BG/BB	−17,007	10.2	10.4	27.7	18.0
Bivariate RF logit	−10,694	11.3	14.9	28.8	47.3
Single death process	−11,539	9.5	11.2	23.5	35.5

expected number of weeks in which each type of transaction is conducted during the forecasting period conditional on the number of weeks with incidences during the calibration period. We present these conditional expectations in Figure 2.

Again, we see that the predictions from our model closely align the number of weeks with purchases in the forecasting period. The independent BG/BB models also track the observed activity during the forecasting period reasonably well, though they overestimate the amount of activity during the forecasting period that is conducted by individuals who were heavy users during the calibration period more so than the proposed model.

In addition to examining the expectations with respect to each activity, we also examine expectations of purchasing given the amount of posting during the calibration period. We present the conditional expectations in Figure 3. Both the BG/BB and our proposed

Figure 1 Weekly Purchasing (a) and Posting (b) Activity

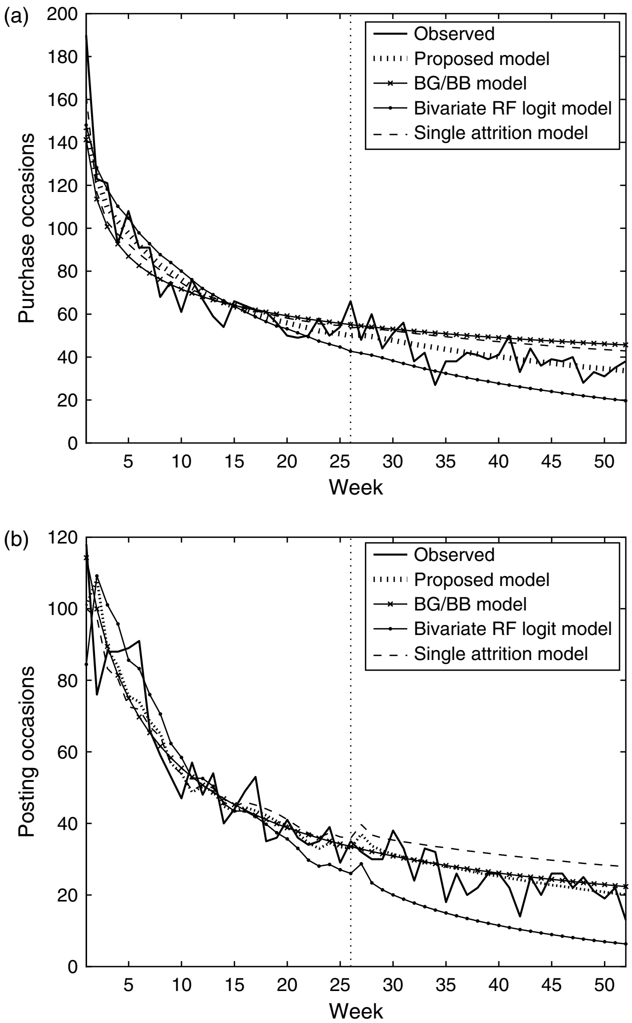


Figure 2 Expected Purchasing (a) and Expected Posting (b) During the Forecasting Period

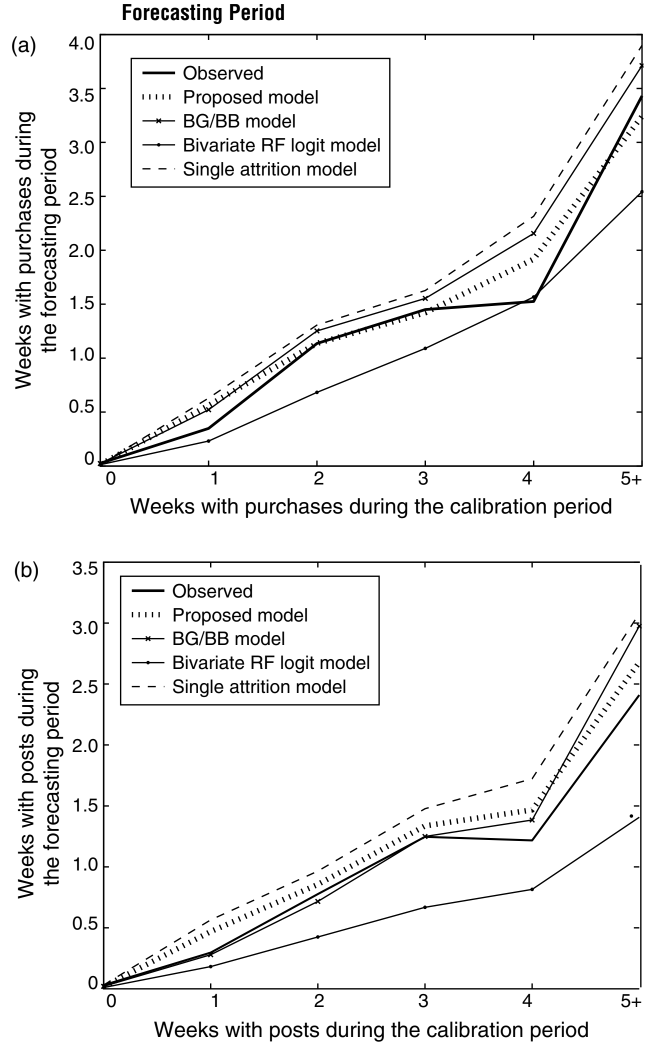


Figure 3 Expected Purchasing During the Forecasting Period Based on Calibration Posting

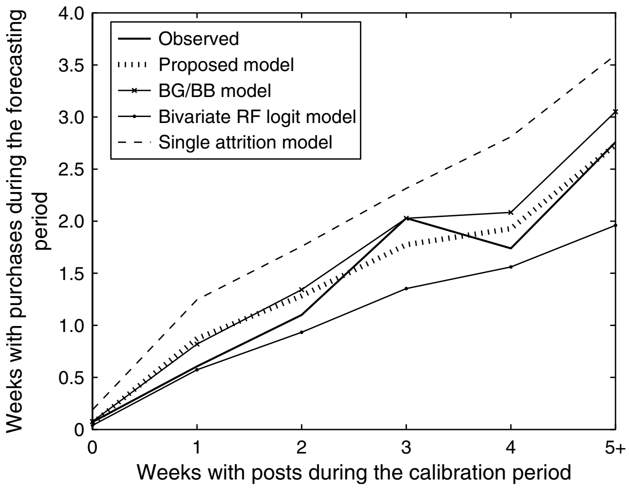
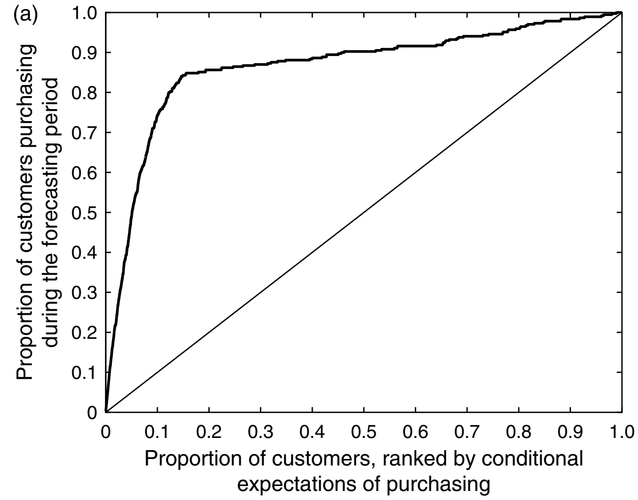
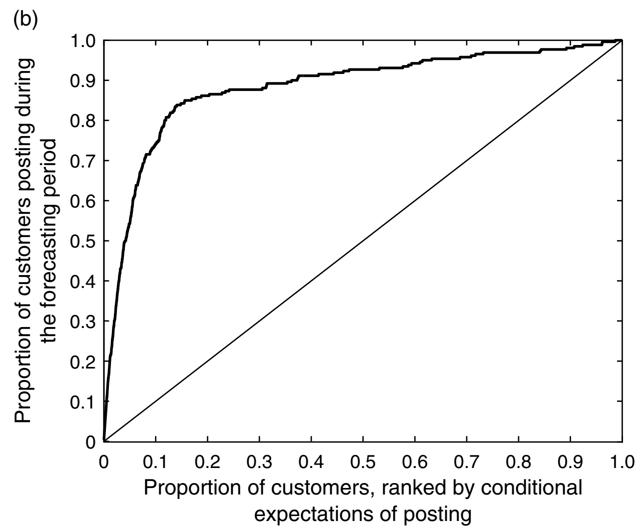


Figure 4 Identifying Likely Purchases from Conditional Purchasing Expectations (a) and Likely Posters from Conditional Posting Expectations (b)





model do reasonably well in capturing the amount of purchasing in the forecasting period, though the proposed model appears to track the observed purchasing activity more closely.

As a further assessment of model performance, using the observed and expected number of weeks with purchases and posts during forecasting period, we calculate the absolute difference between the observed and expected number of weeks with transactions involving each activity, and then we sum this difference across the activities, providing a customer-level error measure. We then calculate the proportion of customers for which the proposed model outperforms each benchmark. We find that the proposed model outperforms each of the benchmarks. It has a lower total absolute error for 61% of customers compared with the single attrition model, 79% of customers compared with the bivariate logit model, and 85% of customers compared with the independent BG/BB models.

Based on our model's predictive ability compared to the benchmark models we consider, we focus the remainder of our discussion on the results from our proposed modeling framework. To illustrate the model's ability to identify those customers who are likely to conduct transactions during the forecasting period, we rank customers in descending order using the expectations of purchasing and posting incidences during the forecasting period. In Figure 4, we present the proportion of customers conducting a transaction during the forecasting period based on these rankings (Schweidel and Knox 2013).

As Figure 4 illustrates, the rankings based on expectations are able to identify those individuals likely to conduct transactions during the forecasting period. Customers in the top 5% of expectations of purchasing based on our model account for 49% of customers who make a purchase during the forecasting period, and the top 10% accounts for 74% of customers purchasing in the forecasting period. Similarly, the top 5% of customers based on expectations of posting account for 54% of customers who post during the forecasting period and the top 10% account for 74% of customers who post.

4.2. Model Inferences

We present the posterior means (and 95% highest probability density (HPD) intervals) of the parameters of interest in Table 4.

As the mean purchasing and posting utilities reflect, we observe a lower incidence of posting compared with purchasing. Based on the decay parameters associated with the stock variables for purchasing (δ_1) and posting (δ_2), the effect of prior activities on subsequent transactional utility diminishes fairly quickly as time passes. Focusing first on the effect of prior activities on purchasing utility, we find that customers are less likely

Table 4 Model Parameters

Parameter	Interpretation	Mean (95% HPD)	$\sqrt{\Sigma}$ (95% HPD) for customer-level effects
α_{h1}	Mean purchasing utility	−2.16 (−2.34, −2.02)	0.76 (0.66, 0.86)
α_{h2}	Mean posting utility	−3.26 (−3.43, −3.06)	0.68 (0.57, 0.80)
β_{11}	Shift in purchasing utility from prior purchasing	−0.83 (−1.00, −0.65)	
β_{12}	Shift in purchasing utility from prior posting	0.00 (−0.19, 0.18)	
β_{21}	Shift in posting utility from prior purchasing	2.04 (1.87, 2.20)	
β_{22}	Shift in posting utility from prior posting	0.26 (0.08, 0.46)	
δ_1	Decay rate for purchasing stock	0.31 (0.23, 0.39)	
δ_2	Decay rate for posting stock	0.25 (0.02, 0.59)	
θ	Coincidence in activity incidence	1.86 (1.70, 2.01)	
κ_{h1}	Mean tendency to become inactive on purchasing	2.12 (1.18, 3.14)	0.89 (0.67, 1.09)
κ_{h2}	Mean tendency to become inactive on posting	−1.83 (−3.43, 0.12)	2.60 (2.10, 3.01)
ψ_{11}	Shift in tendency to become inactive on purchasing from prior purchasing	0.45 (−1.37, 1.94)	
ψ_{12}	Shift in tendency to become inactive on purchasing from prior posting	−2.26 (−4.90, 0.87)	
ψ_{21}	Shift in tendency to become inactive on posting from prior purchasing	1.28 (0.20, 2.40)	
ψ_{22}	Shift in tendency to become inactive on posting from prior posting	−1.93 (−4.37, −0.23)	
τ_1	Shift in tendency to become inactive on purchasing if still active on posting	−6.51 (−7.89, −5.25)	
τ_2	Shift in tendency to become inactive on posting if still active on purchasing	−2.54 (−5.14, −0.34)	
ω	Coincidence in latent attrition processes	−0.55 (−3.64, 2.06)	
π_1	Proportion starting active on purchasing and posting	0.18 (0.15, 0.20)	
π_2	Proportion starting active on purchasing only	0.03 (0.00, 0.06)	
π_3	Proportion beginning active on posting only	0.17 (0.11, 0.24)	

to make a purchase in the weeks subsequent to making a purchase (β_{11}).[6] This finding could be consistent with customers being stocked up on purchasing in the weeks subsequent to a purchase (e.g., Ailawadi and Neslin 1998, Bell et al. 1998), with customers more likely to make another purchase after some time has elapsed. Another potential explanation may have to do with the nature of the product, as customers may be satiated (e.g., Redden 2008) and not have a taste for immediately consuming more of the digital content offered by the website.

Unlike the effect of prior transactions on subsequent purchasing, we see that customers are more likely to post following weeks in which they conducted a purchase (β_{21}) or post (β_{22}), with past purchases exhibiting a larger impact on posting utility. This finding suggests a link between customers' purchasing behavior and contribution of user-generated content when customers are in the relationship state in which they are active on both activities. As we will discuss shortly, we also see a relationship between the attrition processes related to these activities.

[6] The negative impact of prior purchasing activity on future purchases was found to be robust across a number of alternative model specifications, including ones in which a single stock variable was constructed as a weighted sum of the purchasing and posting stock measures and a specification with a single attrition process. We note that this finding may be specific to our empirical context and may not generalize to applications of our modeling framework.

As one might expect, we find positive coincidence between purchasing and posting (θ), indicating that the two activities are likely to occur in the same week. Recall, though, that coincidence between multiple activities only occurs when a customer is still active on these activities. Firms may only take advantage of the synergies between activities for those customers who are still active, which may affect those customers targeted by their promotional efforts. We illustrate this in the next subsection.

Turning our attention to the attrition process, we see that the tendency to become inactive on each activity depends on whether the customer is still active on the other activity. Specifically, when customers are active on both purchasing and posting, they are less likely to become inactive on purchasing (τ_1) or posting (τ_2). This suggests that customer base analysis in which customers may engage in multiple activities needs to move beyond the standard recency, frequency, and monetary metrics and capture the additional "touch-points" that customers generate—in our case, through posting activity (Berger et al. 2002). In accounting for these touchpoints, note that we do not treat them as exogenous; rather, we model the buy 'til you die process for both activities and consider how the different activities in which customers engage are interrelated. Our findings are consistent with Reinartz and Kumar (2003), who find that the number of categories in which customers shop is positively associated with longer relationships and hence reduced customer attrition. Our analysis, however, considers the possibility that attrition may occur activity by activity and that these processes may be related, which has not been investigated.

In contrast to the transaction incidence processes where we find positive coincidence between activities, we do not find evidence of coincidence in the latent attrition processes. Although we do not find evidence of prior transactional activities affecting the tendency to become inactive on purchasing, we do find that prior transactions affect the tendency to become inactive on posting. Whereas prior purchasing has a positive impact on the likelihood of customers becoming inactive with respect to posting (ψ_{21}), the effect of prior posting (ψ_{22}) is negative and larger in magnitude. This suggests that prior posts have both a short-term impact on the likelihood of making another post (β_{22}) and a long-term effect by reducing the likelihood with which customers become inactive on posting. As the tendency to churn with regard to purchasing is lower when customers are still active with regard to posting, this suggests a mechanism through which posting contributes to the revenue generated by posting. Although prior posts do not affect the immediate purchase likelihood immediately, customers engaging in posting are more likely to be active in the future on both activities, which is essential for future purchasing. We next illustrate how firms may take advantage of this relationship to inform how they manage customer relationships.

4.3. Assessing Latent Attrition, Activity by Activity

A common way in which latent attrition models have been employed in customer base analysis is to assess the likelihood that a customer is still alive (e.g., Reinartz and Kumar 2000). In a multiactivity noncontractual setting, this can be done activity by activity to recognize the relationships that exist among the activities. Specifically, we must take into account the linked tendency for attrition on each activity depending on the set of activities that a customer is still active (τ_1 and τ_2). To do so, we calculate the posterior belief of a customer's state membership, taking into account the sequence of activities conducted to date, using Bayes' rule (Netzer et al. 2008, Schweidel et al. 2011). The posterior probability of customer h being in state s at time t is given by

$$\Pr(\text{customer } h \text{ is in state } s \text{ at time } t) = \frac{\boldsymbol{\pi} D_{h1} W_{h1} D_{h2} W_{h2} \cdots W_{h,t-1}(s)\, d_{hts}}{\boldsymbol{\pi} D_{h1} W_{h1} D_{h2} W_{h2} \cdots W_{h,t-1} D_{ht} \mathbf{1}}, \tag{14}$$

where d_{hts} is the sth diagonal element of D_{ht}, the denominator is the likelihood of the observed data through time t, and $W_{h,t-1}(s)$ refers to the sth column of the transition matrix $W_{h,t-1}$.

Table 5 Illustrative Posterior State Membership Probabilities

Activities conducted at $t=12$	Pr(active on both activities)	Pr(active on purchasing only)	Pr(active on posting only)	Pr(inactive)
None	0.54	0.06	0.01	0.39
Purchasing only	0.91	0.09	0.00	0.00
Posting only	0.98	0.00	0.02	0.00
Purchasing and posting	1.00	0.00	0.00	0.00

To illustrate how transactions of one activity may affect beliefs about the likelihood of future transactions of another activity, consider a customer who is first observed to both purchase and post at $t=1$. As such, we know that he begins active on both activities. We assume that he conducts no activity until $t=12$. We consider four different scenarios, depicting the possible combinations of purchasing and posting incidence that may occur. For each scenario, we calculate the posterior state membership (averaged across iterations), which is shown in Table 5.

When the customer purchases and posts at $t=12$, it is known that he is active on both activities. If the customer only posts at $t=12$, it is most likely that the customer is still active on both activities, as he has a decreased tendency to become inactive on purchasing while still active on posting (τ_1). As a result, the additional touchpoint of the posting activity provides a valuable signal as to whether the customer may purchase in the future. Even though we have not observed a purchase since $t=1$, he still likely has the potential to purchase in the future. Meanwhile, when neither purchasing nor posting are observed in $t=12$, there is a 39% chance that he is already inactive on both activities.

As we show in Table 5, the additional touchpoint provided by posting activity allows us to differentiate between customers who may still purchase in the future and those who have more likely ceased their relationship with the firm. In contexts in which there is a single activity of interest, models such as the BG/BB can be employed to forecast customers' future activities, score customers, and allocate resources accordingly. When multiple transactional activities are observed, such as purchasing and the posting of digital content as in our empirical context, both streams of transactional activity can be used in these tasks. As we have shown, this has the potential to more accurately forecast future activity, which can enable firms to more efficiently deploy its marketing efforts.

Firms managing customer relationships can use such information to allocate resources more effectively across prospective and future customers. This may entail increasing marketing efforts directed at some customers (or prospects) and decreasing efforts focused on others. Such efforts should be evaluated in terms of the impact they will have on customers' future revenue-generating efforts. With data on the firm's marketing efforts, our modeling framework could be employed to estimate the net effect of marketing activity on customers' residual value in a multiactivity transactional setting, recognizing that marketing aimed at promoting one activity may benefit not only that specific activity but other activities offered by the firm.

5. Conclusions and Future Research

Using the data on purchasing and posting activity at an e-commerce website, we generalize customer base analysis to a multiactivity transactional setting. Our model offers superior performance compared with benchmarks by allowing for the latent attrition processes and the incidence processes while customers are still active to be related to each other. We examine how the latent attrition processes of multiple activities that customers conduct may be linked, which to the best of our knowledge has not previously been explored. We consider a number of ways in which multiple activities may be associated by incorporating correlated model parameters across activities, coincidence between activities in both the incidence processes and latent attrition processes, the impact of recent activity use on subsequent transactions and latent attrition, and a dependency between the tendency for attrition on one activity and whether a customer is still active on another activity. Though subsets of these sources of association between activities have been considered in extant research (e.g., Manchanda et al. 1999, Park and Fader 2004, Niraj et al. 2008, Schweidel et al. 2011), they have not been examined jointly.

Although it is an empirical question as to how the latent attrition processes for multiple activities may be related to each other, it is a critical element to managing customer relationships, as it will affect organizations' abilities to cross sell offerings to its customers (e.g., Kamakura et al. 2003, Li et al. 2011). We find that the processes are linked, suggesting that customers are less likely to become inactive on one activity while they are active on other activities. We also find that customers' decisions to remain active with respect to posting digital content (and hence making it available to others) are affected by prior transactions. As we demonstrate, these relationships affect inferences about the likelihood with which customers may conduct activities in the future. By incorporating touchpoints beyond purchasing into our analysis, we can better distinguish those customers who are likely inactive from those who may make purchases in the future and thereby provide guidance for a firm's resource allocation across customers (e.g., Schweidel and Knox 2013). Our research also suggests the potential value associated with encouraging the production and posting of user-generated content (e.g., Moe and Trusov 2011, Moe and Schweidel 2012), as it may contribute to revenue by reducing the likelihood with which customers cease conducting purchases.

A number of avenues remain open for future research on both managerial and methodological fronts. One direction in which this stream of research can continue is to examine the return on marketing associated with encouraging different types of activities in which customers engage. For example, if a price promotion were to directly contribute to increased purchasing while encouraging the contribution of content for viewing by others could also contribute to increased purchasing, a firm may seek to identify the amount of marketing whose impact is equivalent to a given price promotion (e.g., Arora and Henderson 2007). In doing so, the firm may quantify the value associated with user-generated content as a promotional tool, weighing the benefits and costs associated with encouraging customers to engage in social media activities.

Our analysis of two types of activities allows us to examine the relationship between latent attrition processes of the activities, the main focus of our modeling effort. We have strived to maintain a parsimonious model with a structure that could be generalized to additional activities, such as for multibrand retailers. In doing so, future research may consider model structures such as coupled or factorial HMMs (e.g., Ghahramani and Jordan 1997, Rezek et al. 2002). One key consideration in pursuing such research will be capturing the associations that exist among multiple activities while avoiding the "curse of dimensionality" (e.g., Park et al. 2014).

Although we have considered customers' choice behavior, it may be worthwhile to simultaneously investigate the volume of activity conducted in each transaction. In doing so, it is important to consider the associations that may exist between the volume and frequency of transactions, as well as with the latent attrition process. The use of copulas (e.g., Danaher and Smith 2011) may offer a means of accounting for such associations.

Acknowledgments

The authors thank Pete Fader and seminar participants at Hewlett-Packard Labs and 2010 Marketing Science Conference for their comments. The authors also thank a company that wishes to remain anonymous for providing the data used in this study. This research was done while the second author was visiting New York University. He is grateful for the generous and cooperative support that he received from the Department of Marketing at New York University's Stern School of Business.

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.

Ailawadi KL, Neslin SA (1998) The effect of promotion on consumption: Buying more and consuming it faster. *J. Marketing Res.* 35(3):390–398.

Arora N, Henderson T (2007) Embedded premium promotion: Why it works and how to make it more effective. *Marketing Sci.* 26(4):514–531.

Bell DR, Ho T-H, Tang CS (1998) Determining where to shop: Fixed and variable costs of shopping. *J. Marketing Res.* 35(3):352–369.

Berger PD, Bolton RN, Bowman D, Briggs E, Kumar V, Parasuraman A, Terry C (2002) Marketing actions and the value of customer assets: A framework for customer asset management. *J. Service Res.* 5(1):39–54.

Cox DR (1972) The analysis of multivariate binary data. *J. Roy. Statist. Soc. Ser. C* 21(2):113–120.

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

Danaher PJ, Smith MS (2011) Modeling multivariate distributions using copulas: Applications in marketing. *Marketing Sci.* 30(1): 4–21.

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

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.

Fader PS, Hardie BGS, Shang J (2010) Customer-base analysis in a discrete-time noncontractual setting. *Marketing Sci.* 26(6):1086–1108.

Ghahramani Z, Jordan MI (1997) Factorial hidden Markov models. *Machine Learn.* 29(2–3):245–273.

Kamakura WA, Wedel M, de Rose F, Mazzon JA (2003) Cross-selling through database marketing: A mixed data factor analyzer for data augmentation and prediction. *Internat. J. Res. Marketing* 20(1):45–65.

Li S, Sun B, Montgomery AL (2011) Cross-selling the right product to the right customer at the right time. *J. Marketing Res.* 48(4): 683–700.

MacDonald IL, Zucchini W (1997) *Hidden Markov and Other Models for Discrete-Valued Time Series* (Chapman & Hall, New York).

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

Moe WW, Schweidel DA (2012) Online product opinions: Incidence, evaluation, and evolution. *Marketing Sci.* 31(3):372–386.

Moe WW, Trusov M (2011) The value of social dynamics in online product ratings forums. *J. Marketing Res.* 49(3):444–456.

Neslin SA, Rhoads EE, Wolfson P (2009) A model and empirical analysis of patient compliance and persistence in pharmaceuticals. Working paper, Dartmouth College, Hanover, NH.

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

Niraj R, Padmanabhan V, Seetharaman PB (2008) A cross-category model of households' incidence and quantity decisions. *Marketing Sci.* 27(2):225–235.

Park CH, Park Y-H, Schweidel DA (2014) A multi-category customer base analysis. *Internat. J. Res. Marketing*. Forthcoming.

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

Redden JP (2008) Reducing satiation: The role of categorization level. *J. Consumer Res.* 34(5):624–634.

Reinartz WJ, Kumar V (2000) On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing. *J. Marketing* 64(4):17–35.

Reinartz WJ, Kumar V (2003) The impact of customer relationship characteristics on profitable lifetime duration. *J. Marketing* 67(1):77–99.

Rezek I, Gibbs M, Roberts SJ (2002) Maximum a posteriori estimation of coupled hidden Markov models. *J. VLSI Signal Processing* 32(1–2):55–66.

Russell GJ, Petersen A (2000) Analysis of cross category dependence in market basket selection. *J. Retailing* 76(3):367–392.

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.

Schweidel DA, Knox G (2013) Incorporating direct marketing activity into latent attrition models. *Marketing Sci.* 32(3):471–487.

Schweidel DA, Bradlow ET, Fader PS (2011) Portfolio dynamics for customers of a multiservice provider. *Management Sci.* 57(3): 471–486.

Singh S, Borle S, Jain D (2009) A generalized framework for estimating customer lifetime value when customer lifetimes are not observed. *Quant. Marketing Econom.* 7(2):181–205.

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.