Available online at www.sciencedirect.com
ScienceDirect
Journal of Interactive Marketing 36 (2016) 77 – 90
JOURNAL OF INTERACTIVE MARKETING
www.elsevier.com/locate/intmar

The Effects of Channel Experiences and Direct Marketing on Customer Retention in Multichannel Settings☆

Chun-Wei Chang [a,*] & Jonathan Z. Zhang [b]

[a] College of Business, Governors State University, University Park, IL 60484, United States
[b] Foster School of Business, University of Washington, Seattle, WA 98195, United States

Abstract

In customer relationship management (CRM), it is critical for managers to understand how and when customers terminate their relationships with the company in order to make more accurate predictions for CLV. However, in many non-contractual settings, customer churn is not easily observed, which presents difficulty for estimating customer retention. In this research, we present a framework for estimating multichannel customer relationship dynamics in a non-contractual setting that flexibly allows for relationship revival and investigates the effects of different channel experiences and marketing communication on retention and profitability. We use a multi-segment, multivariate hidden Markov modeling framework to model three managerially relevant customer behaviors: purchase amount, purchase incidence, and channel choice. Using data from a multichannel clothing retailer, we uncover two latent relationship states that customers migrate to and from — an active state and an inactive state characterized by different levels of purchase frequency, responsiveness to marketing, and profitability. We find that an offline (retail-store) channel can be used to migrate customers from an inactive state to an active state, effectively serving the purpose of “education” or “revival,” whereas an online channel is most effective in keeping the existing active customers active, thus serving the purpose of “retention”. Using counterfactual analysis, we highlight an opportunity for the multichannel firm to optimize marketing strategies to dynamically manage and increase the retention and hence also the value of its customer base.

Keywords: Customer retention; Multichannel marketing; Hidden Markov model; Customer–firm relationship

Introduction

The multichannel marketing environment is becoming increasingly prevalent in recent years. Firms and their customers can interact via brick-and-mortar stores, catalogs, online stores, emails and in recent years mobile platforms, and multichannel marketing has become an important tool to motivate customers to shop more frequently through increased interaction and to build lasting customer relationships (Hansotia and Rukstales 2002; Rangaswamy and Van Bruggen 2005). In addition, multichannel firms are looking for strategies to increase customer retention and avoid customer churn, as the costs of customer acquisition are much higher than that of retention, and small increases in retention could drive large profit increases (Pfeifer and Farris 2004; Reichheld and Sasser 1990). Specifically, Gupta and Lehmann (2003) assert that an increase in retention of 5% yields a dramatic 22% to 37% increase in customer lifetime value.

Given the prevalence of the multichannel environment and the importance of customer retention, it is crucial for both marketing academics and practitioners to study the link between these two areas in order to answer the question of how to increase customer retention and thus increase customer value in multichannel settings. However, such research is particularly difficult for firms in non-contractual settings (e.g. most retail settings such as Nordstrom, Sephora, L.L. Bean, and the recently opened Amazon stores), because the termination of relationships is difficult to observe, and thus retention rates

☆ The authors contributed equally, and are listed alphabetically.
* Corresponding author at: College of Business, Governors State University, University Park, IL 60484, United States.
E-mail addresses: cchang@govst.edu (C.-W. Chang), zaozao@uw.edu (J.Z. Zhang).

http://dx.doi.org/10.1016/j.intmar.2016.05.002
1094-9968/© 2016 Direct Marketing Educational Foundation, Inc., dba Marketing EDGE.

cannot be easily evaluated. Our research framework studies customer retention in multichannel settings, while accounting for issues in non-contractual settings.

Most extant research on customer retention has either focused on highlighting the importance of retention on customer lifetime value (Gupta and Lehmann 2003; Pfeifer and Farris 2004; Reichheld and Sasser 1990) or on attempting to estimate the impacts of various factors on relationship length in contractual settings that require accurate information about the termination of the relationship (Boehm 2008; Schweidel, Fader, and Bradlow 2008). From a channel perspective, some studies have examined the impact of Internet use (specifically, online banking) on service termination (customer retention) (Boehm 2008; Campbell and Frei 2010; Hitt and Frei 2002; Verhoef and Donkers 2005). Those studies are also based on contractual settings.

These issues are hard to resolve in non-contractual settings due to the difficulty in identifying a customer's true relationship with a firm. The extant literature on customer relationships suggests that customers are considered active or inactive based on their purchase activities, which are governed by underlying relationship states. These latent relationship states that govern customer activity could exhibit different levels of awareness or top-of-mind awareness, different levels of trust towards the retailer, or different levels of familiarity (Li, Sun, and Montgomery 2011; Montoya, Horton, and Kirchner 2008; Zhang, Netzer, and Ansari 2014), all of which could change over time via repeated interactions with the firm. Retention efforts in a non-contractual setting should be aimed at preventing active customers from becoming inactive and reviving inactive customers back to an active state. However, in non-contractual settings, customers may seem to have churned when they have not actually terminated the relationship with a firm; customers with longer inter-purchase time might just be in an inactive relationship state and need to be distinguished from customers who have truly terminated the relationship altogether. Previous research that attempts to derive customer retention with a non-contractual firm 1) has treated all purchases the same and has not accounted for the channels through which customers make purchases, ignoring the fact that different channels offer different experiences (i.e. retail stores offer richer experience, online offers convenience) and would impact the relationship differently, and 2) has not considered customer revival. In particular, the Pareto/NBD and BG/NBD models, which explore repeat purchase behavior while accounting for unobserved customer dropout rates in CRM, ignore the possibility that customers who shop through a particular channel may have different retention/churn probabilities than those who shop through alternative channels. These models treat the inactive customer as "dead" and do not allow for the customer to come back to the active relationship state. Overall, the various impacts of multichannel experiences and direct marketing on customer retention in non-contractual settings have not been well explored.

The purpose of this research is to fill such a gap and to investigate the impacts of multichannel experiences and direct marketing on improving customer–firm relationships and customer retention in non-contractual settings. To do so, we simultaneously model three managerially important customer decisions for a multichannel retailer – purchase incidences, channel choices and purchase amounts – and recover the underlying latent relationship states which represent customers' tendencies to stay active or inactive. We apply a multi-segment, multivariate hidden Markov Model (HMM) approach to link the latent relationship states with the three observed customer behaviors (i.e. incidence, channel choice, and amount) and examine how choice preference evolves as a customer–firm relationship changes due to channel experiences and marketing communications. This modeling framework allows us to first identify each customer's latent state of relationship state over time; second, to determine the retention probability affected by channel experience; third, to identify customers who are more likely to churn by the end of the observation period; and finally, to examine the impact of alternative channel experiences and direct marketing on migrating customers towards desirable state, hence improving retention.

We contribute to the CRM and multichannel literature in the following dimensions. Methodologically, our framework investigates the effects of channels and direct marketing on customer behaviors in a non-contractual setting through multiple important customer decisions. It can track each customer's varying retention rates dynamically and flexibly allows for customer relationships to be revived, a framework that is relevant for many multichannel markers. Theoretically, to the best of our knowledge, our paper is among the first to provide empirical evidence for the different purposes of channels in managing customer portfolios in non-contractual settings: offline retail channel can be used to migrate customers from an inactive to an active state, effectively serving an "educational" or "revival" purpose, whereas online channels have the biggest impact of keeping currently active customers active, thus serving a "retention" purpose. These findings are consistent with various prior research that suggests that offline channels are experientially more immersive (Verhoef, Neslin, and Vroomen 2007), which can help get customers excited about the retailer experience and hence is good for learning and relationship building, whereas online channels provide convenience and low transactional cost for those customers who are familiar with the retailer experiences (Lal and Sarvary 1999). The catalog channel's effect is weak compared to both the offline and online channels, which confirms the intuition that it has neither the immersive experience of a retail immersive experience of a retail store nor the convenience of an online channel and serves to explain why the catalog is a "sunset" channel that's losing its appeal. Managerially, our framework allows the firm to identify every customer's latent state every month, which enables the firm to adjust its marketing resources dynamically for each customer. Recent multichannel research has demonstrated that customers can be persuaded through non-financial marketing communication to adopt a particular channel, suggesting right channeling as a marketing strategy that firms could employ (Montaguti, Neslin, and Valentini 2016). Using the counterfactual analyses of both customer activity and profitability, our findings on the roles of channels and marketing provide managerial guidance on how the firm can use marketing strategies to dynamically manage its customer base and increase its value.

In the following section, we review issues related to customer retention and HMM as applied in the field of CRM. In the Model Development section, we discuss the structure of our HMM, followed by the Empirical Application section, which describes the empirical application of the proposed model using panel data from a large multichannel clothing retailer. The Counterfactual Analysis section provides a counterfactual analysis that highlights the opportunities to use marketing to manage customer–firm relationships, improve retention, and increase the value of the customer base. In the Conclusions and Directions for Future Research section, we discuss theoretical and practical contributions and conclude with directions for future research.

Applying HMM to Multi-channel Customer Retention

Customer Retention in Non-contractual Settings

Retention in many works of CRM literature refers to a single and constant ratio used to represent the portion of retained customers and to calculate lifetime value, which means that the estimated retention probabilities do not vary over a customer's lifetime (Blattberg and Deighton 1996). Many of the research studies assert that small increases in retention drive large increases in profits (Gupta and Lehmann 2003; Pfeifer and Farris 2004; Reichheld and Sasser 1990), thus inaccurate estimates of retention rates would lead to large biases for the value of a customer base. Most previous research in contractual settings uses a family of hazard models to address retention duration, predict the customer's lifetime, and examine the impact of predictors on relationship length (Boehm 2008; Fader and Hardie 2010; Schweidel, Fader, and Bradlow 2008). While these approaches are appropriate in contractual settings with clearly stated drop-out times, they could not be readily applied in non-contractual settings where customers' retention tendencies are unobserved and where relationships can change over time, thereby making the retention rate time varying.

In non-contractual settings, issues around how to estimate customer lifetime value based on accurate retention and how to count retained customers persisted until the Pareto/NBD (Reinartz and Kumar 2000; Schmittlein, Morrison, and Colombo 1987) and BG/NBD (Fader, Hardie, and Lee 2005) models were proposed. In the context of CRM, the Pareto/NBD and BG/NBD models explore such issues as predicting future demand, customer churn, and retention rate by assuming that customers may transition from an "active" state to an "inactive" state at different rates. The two models attempt to estimate customer retention and dropout rate with slightly different assumptions and provide good answers to questions about how many customers will be active or "alive" in the future given their past behavior. One major property of the Pareto/NBD and BG/NBD models is that both assume that customers start off as active and, based on observed activities, will experience a discreet jump to an inactive or "dead" state at some point and do not switch back to an active state. Thus, they imply that there is no spectrum between active and inactive and that customers who are identified as inactive remain inactive permanently. This is a strong restriction on estimating retention and neglects important factors such that relationships are inherently gradual and dynamic in nature and precludes the possibility that even inactive relationships might be revived through marketing interventions. In our modeling framework, we model latent relationships as a gradual process instead of discreet jumps, and flexibly allows for the possibility that the customer could be revived.

Beyond just the retention rate number, managers are interested in understanding the factors underlying retention rates; for example, under which circumstances would a customer come back after being inactive and under which circumstances retention would increase. Also, as different channels have different value propositions, customers who are used to purchasing through one channel may have a different relationship and hence a different retention rate than those who purchase through an alternative channel. Therefore, the effect of channel experiences on retention should be considered. Our HMM framework not only allows for customers to evolve flexibly among relationship states, but also explicitly models the impact of channel experiences and marketing on retention by incorporating those factors into customer state transitions.

Table 1 summarizes the relevant extant research along the above important dimensions and highlights our study's comprehensiveness and contributions.

HMM Applications and Customer Relationship Management

Hidden Markov models are well suited for our research tasks and have been used recently to infer latent relationship states from observed transactions, such that customers can flexibly migrate between states (Du and Kamakura 2006; Luo and Kumar 2013; Montoya, Netzer, and Jedidi 2010; Netzer, Lattin, and Srinivasan 2008; Schweidel, Bradlow, and Fader 2011; Zhang, Netzer, and Ansari 2014). Using data on alumni donation behaviors, Netzer, Lattin, and Srinivasan (2008) assign donors to active or inactive latent states according to whether they express low, medium, or high interest in donating; Schweidel, Bradlow, and Fader (2011) consider latent states pertaining to customers' propensity for service usage with a multiservice provider; and Li, Sun, and Montgomery (2011) use purchase data on various financial products to model states as consumers' latent financial sophistication.

The merits of employing HMM to study non-contractual customer relationships and retention are flexibility and parsimony. Unlike the Pareto/NBD and BG/NBD models, HMM does not impose a priori constraints on the number of states and transition paths; the total number of states is instead determined empirically through model selection criteria. Beyond the flexibility to identify the number of states empirically, HMM can show how "transient" or "sticky" different states are, allowing for gradual migrations in relationship states and allowing for transition from one state to all other states. These properties are well suited for customer–firm relationships, which take time to either develop or decay. Whereas models relying on observed state variables such as previous purchase amount or

Table 1
Comparing the proposed multichannel retention framework with existing models.

Type of model	Representative research	Multi-channel	Contractual vs. non-contractual	Model behavior	Allows for retention rates to vary among customers	Allows for relationship revival
CLV	Blattberg and Deighton (1996)	No	Contractual	CLV	No (aggregate retention rate)	No
	Gupta and Lehmann (2003)	No	Contractual	CLV	No (aggregate retention rate)	No
	Pfeifer and Farris (2004)	No	Not specific (general numerical illustration by sensitive analysis)	CLV	No (aggregate retention rate)	No
Hazard model	Boehm (2008)	No	Contractual	Relationship length	Yes	No
Proportional hazards model	Schweidel, Fader, and Bradlow (2008)	No	Contractual	Service retention	Yes	No
Shifted-beta-geometric (SBG)	Fader and Hardie (2010)	No	Contractual	Retention	Cohort level	No
Pareto/NBD	Reinartz and Kumar (2000)	No	Non-contractual	Repeat buying - inter-purchase time	Yes	No
	Schmittlein, Morrison, and Colombo (1987)	No	Non-contractual	Repeat buying - inter-purchase time	Yes	No
BG/NBD	Fader, Hardie, and Lee (2005)	No	Non-contractual	Repeat buying - inter-purchase time	Yes	No
HMM	Netzer, Lattin, and Srinivasan (2008)	No	Non-contractual	Purchase (donation) incidence	Yes	Yes
Multivariate HMM	This paper	Yes (retail store, online, catalog)	Non-contractual	Purchase incidence, channel choice, purchase amount	Yes	Yes

previous channel choice might indicate discreet state jumps from one purchase occasion to the next and lead to overstated movements, HMM can smooth out such discreet shifts. In terms of parsimony, the current state depends only on the previous state and is independent from all previous migration paths. The relationship states that empirically emerge would already contain all relevant information regardless of the customers' past history. Thus, from a practical perspective, managers would only need to focus on the present and not need to keep track of the path the customer took to get to the current state.

In our modeling specification, we allow different channel experiences and marketing communication to influence both the immediate behaviors as well as state transitions, thus allowing us to assess both the short and long-term impact of these factors.

Model Development

Customers who are accustomed to shopping in a single channel may have different retention probabilities than customers who shop in multiple channels. As the previous literature asserts, multichannel shoppers are more active (Kumar and Venkatesan 2005), exhibit lower churn rates, and demonstrate higher propensity to buy more (Stone, Hobbs, and Khaleeli 2002). Also, different purchasing patterns across channels should be accounted for; e.g., a customer who makes seven purchases in a retail store and three purchases online should be distinguished from a customer who makes three purchases in a retail store and seven purchases online, although both situations result in a total of ten purchases. Our proposed model deals with these issues by explicitly modeling purchase activity and channel choice as a function of various channel experiences.

Given the relationship state, the customer makes observed purchase decisions, namely, when to buy, how much to buy, and which channel to buy from. The relationship states are unobserved but can vary with respect to the observed customer decisions. The HMM can account for the varying impacts of channel experiences and marketing communication on customer retention and link these latent relationship states with observed decisions. Furthermore, we overlay a latent-class specification on top of the HMM to account for time-invariant customer heterogeneity apart from customer dynamics, thus resulting in a multi-segment multivariate HMM.

The HMM has 4 components: 1) The state-dependent choice distribution; 2) the Markov chain transition matrix among states that explains how the customer moves from one period to the next, as well as the effects of channel experiences and marketing on transition; 3) the initial state distribution, which denotes a customer's state at the beginning of the observation period; and 4) the customer's latent state probability in each time period. We now describe them in detail in the following subsections.

State-dependent Choice Distribution

Customers are involved in three decisions during each period: whether to make a purchase, how much to purchase, and where to purchase (e.g., retail store, online, or catalog). The three choices of purchase incidence, channel selection, and purchase amount depend on a customer's relationship state with the firm. The choices the *nth* customer makes in time period t given state $S_{nt}=i$ are defined as follows.

$$Y_{nt}|(S_{nt}=i) = Y_{nt|i} = \left(B_{nt|i}, C_{nt|i}, Q_{nt|i}\right),$$

where $B_{nt|i}$ equals 1 if the nth customer makes a purchase, and zero otherwise. $C_{nt|i}$ is the channel choice, and $Q_{nt|i}$ is the purchase amount. As channel choice can only happen after purchase incidence, we model the incidence and channel choice with the nested multinomial logit structure to satisfy IIA assumption and model the amount as distributed lognormal conditional on purchase incidence. The nested structure can be divided into two parts: purchase probability $P(B_{nt|i})$ and channel choice and amount conditional on purchase $P(Q_{nt|i}, C_{nt|i}|B_{nt|i})$. We let channel choice and amount to be conditionally independent. Therefore, the joint probability of purchase incidence and channel selection can be represented as follows.

$$P\left(Y_{nt|i}\right) = P(Y_{nt}|S_{nt} = i) = P\left(Q_{nt|i},\ C_{nt|i}|B_{nt|i}\right)P\left(B_{nt|i}\right).$$

Conditional Channel Utility

Consider first the channel choice conditional on the purchase. The random utility of channel choice, which includes the deterministic and random components, is

$$U^{v}_{nt|i} = \alpha_{iv} + X_{nt}\beta_{iv} + \varepsilon_{iv}, v\in\{1, \cdots, C\}, \tag{1}$$

and it is specified as a multinomial logit model by utility maximization as follows.

$$P\left(C_{nt|i} = v|B_{nt|i}\right) = \frac{\exp(\alpha_{iv} + X_{nt}\beta_{iv})}{\sum_{v=1}^{C}\exp(\alpha_{iv} + X_{nt}\beta_{iv})}, i = 1, \cdots, m, \tag{2}$$

where α_{iv} is the state-specific intrinsic utility of channel v in state i, $\forall v\in\{1, \cdots, C\}$, X_{nt} is a vector of explanatory variables that are common across channels for customer n at t, and β_{iv} is a vector of the state-and-channel-specific coefficient of variables X_{nt} for channel v in state i. The term $P(C_{nt|i}=v|B_{nt|i})$ represents the probability that customer n chooses channel v given state i at t while making a purchase.

Purchase Amount

We assume the purchase amounts follow a lognormal distribution with p.d.f. and corresponding c.d.f., given by

$$f(Q_{nt}) = \frac{\varnothing\left(\frac{\log(Q_{nt})-W'_{nt}\boldsymbol{\beta}_{\mathbf{(Q)i}}}{\sigma_Q}\right)}{\sigma_Q\ (Q_{nt})},\quad F(Q_{nt}) = \Phi\left(\frac{\log(Q_{nt})-W'_{nt}\boldsymbol{\beta}_{\mathbf{(Q)i}}}{\sigma_Q}\right), \tag{3}$$

where $\boldsymbol{\beta}_{\mathbf{(Q)i}}$ is a vector of state-specific coefficients for covariates W_{nt} that affect the purchase amount in state i. σ_Q is the scale parameter, and $\varnothing$ and Φ represent the p.d.f. and c.d.f. of the standard normal distribution, respectively.

Purchase Incidence

Next, consider the purchase probability of customer n at time t in state i, $P(B_{nt|i})$. Assume that the utility of purchasing R for customer n at t in state i is as follows.

$$R_{nt|i} = \delta_i + Z_{nt}\gamma_i + e_{nt}, \tag{4}$$

where δ_i and γ_i are unknown parameters and Z_{nt} is a vector of covariates, which contribute to the purchase decision incidence. A customer will make a purchase if and only if the maximum channel utility is greater than his utility of not purchasing. The inclusive value for purchasing which is the expected maximum utility of making a purchase at t in state i, is defined as follows.[1]

$$CV_{nt|i} = \ \ln\sum_{v=1}^{C}\exp\left(U^{v}_{nt|i}\right). \tag{5}$$

Therefore, the purchase probability is

$$P\left(B_{nt|i} = 1\right) = \frac{\exp\left(\delta_i + Z_{nt}\gamma_i + CV_{nt|i}\xi_i\right)}{1 + \exp\left(\delta_i + Z_{nt}\gamma_i + CV_{nt|i}\xi_i\right)}\ , 0 \le \xi \le 1, \tag{6}$$

where the parameter ξ_i is restricted to be one to get the non-nested model[2] and unrestricted to allow some degree of heteroscedasticity.

Combining the three components, the state-dependent choice function for customer n at shopping occasion t in state i is therefore:

$$P\left(Y_{nt|i}\right)=\left[1-P\left(B_{nt|i}=1\right)\right]^{\left(1-\omega^{b}_{nt}\right)}\times \left[\prod_{v=1}^{C}P\left(C_{nt|i}=v|B_{nt|i}=1\right)^{\omega^{v}_{nt}}\times P\left(Q_{nt|i}=x|B_{nt|i}=1\right)\times P\left(B_{nt|i}=1\right)\right]^{w^{b}_{nt}}, \tag{7}$$

where $\omega^{b}_{nt} = \begin{cases} 1, & \text{if a customer } n \text{ makes a purchase at } t \\ 0, & \text{otherwise} \end{cases}$,

$\omega^{v}_{nt} = \begin{cases} 1, & \text{if a customer } n \text{ makes a purchase through channel } v \text{ at } t \\ 0, & \text{otherwise} \end{cases}$.

For the purpose of state identification, we restrict the intrinsic purchase probability (δ_i) to be non-decreasing, thus, $\delta_1 \le \delta_2 \le \cdots \le \delta_m$.

The time-varying covariates for customer n at t consist of variables that have an immediate impact on a customer's decision and are discussed in detail in the Empirical Application section.

Markov Chain Transition Matrix

The HMM does not a priori restrict the number of states. Also, it allows for customers staying in a current state or moving to any other state, by estimating a full and flexible transition matrix. Given m states, we assume that the transition matrix $Q(i_{t-1}, i_t)$ is defined as follows.

$$Q(i_{t-1}, i_t) = \begin{bmatrix} q_{11}\ q_{12}\cdots q_{1m-1}\ q_{1m} \\ q_{21}\ q_{22}\cdots q_{2m-1}\ q_{2m} \\ \vdots\quad \vdots\quad \cdots\quad \vdots\quad \vdots \\ q_{m1}\ q_{m2}\cdots q_{mm-1}\ q_{mm} \end{bmatrix}, \tag{8}$$

[1] Greene (2003).

[2] With ξ_i=1, Eq. Eq. (5) reverts to a basic multinomial logit model (Greene 2003) which meets IIA assumption.

where $q_{jk}=P(i_t=k|i_{t-1}=j)$ denotes the transition probability from state j at $t-1$ to state k at t, and $\sum_{k=1}^{m} q_{jk}=1,\ 0\le q_{jk}\le 1$ for all $j,k=1,\cdots,m$.

We model the transition matrix as an ordered logit model (Greene 2003), where the latent relationship states are rank ordered by the propensity to move from inactive to active. This is in contrast with the multinomial logit approach that does not assume a rank-ordered relationship between states. Theoretically, the ordered logit assumption is appropriate in the CRM setting where customers' activities vary on a simple spectrum. Methodologically, estimating ordered logit is more parsimonious than multinomial logit and can save on parameters.

The elements in the transition matrix $Q(i_{t-1},i_t)$ can be defined as follows:

$$q_{j,1}=\frac{\exp\left(\mu_j^1-\theta_j A_{nt}\right)}{1+\exp\left(\mu_j^1-\theta_j A_{nt}\right)}, \tag{9}$$

$$q_{j,k}=\frac{\exp\left(\mu_j^k-\theta_j A_{nt}\right)}{1+\exp\left(\mu_j^k-\theta_j A_{nt}\right)}-\frac{\exp\left(\mu_j^{k-1}-\theta_j A_{nt}\right)}{1+\exp\left(\mu_j^{k-1}-\theta_j A_{nt}\right)}, \tag{10}$$

$$q_{j,m}=1-\frac{\exp\left(\mu_j^{m-1}-\theta_j A_{nt}\right)}{1+\exp\left(\mu_j^{m-1}-\theta_j A_{nt}\right)}, \tag{11}$$

for $j\in\{1,\cdots,m\}$, $k\in\{2,\cdots,m-1\}$, $\mu_j^1<\mu_j^2<\cdots<\mu_j^{m-1}<\mu_j^m$, where θ_j is a vector of parameters on transitions from state j and A_{nt} is the vector of time-varying covariates for customer n between time $t-1$ and time t that impacts the transition and thus has a long-term impact (in our model specification, we use marketing and experience). μ_j^k is the threshold value to a more active state ($k\ge j$) or a more inactive state ($k<j$) for a customer in state j.

Initial State Distribution

Define π_n as a vector of initial probabilities for a customer n ($\pi_n=(\pi_{n1},\pi_{n2},\cdots,\pi_{nm})'$), and the initial state distribution is defined as the stationary distribution of the transition matrix (Netzer, Lattin, and Srinivasan 2008). The initial state distribution is calculated by solving the following equation.

$$\pi_n=\pi_n Q_n,\ \sum_{i=1}^{m}\pi_{ni}=1, \tag{12}$$

where Q_n is the transition matrix.

The HMM Likelihood Function

The customer makes the three decisions conditional on being in state i at time t. These decisions are interrelated as they all depend on the customer's latent state. Y_{nt} is the sequence of the observed combination of purchase incidences, channel choices, and purchase amount (B_{nt}, C_{nt}, Q_{nt}) for a customer n, and S_{nt} is the set of latent relationship states. The joint probability of an observed sequence of choices Y is given by summing over all possible states over time, as follows.

$$\begin{aligned}L_{nT}&=P(Y_{n1}=y_{n1},\ Y_{n2}=y_{n2},\cdots,Y_{nT}=y_{nT})\\&=\sum_{\{S\}}P(Y_{n1}=y_{n1},Y_{n2}=y_{n2},\cdots,Y_{nT}=y_{nT}|S_{n1}=i_1,S_{n2}\\&=i_2,\cdots,S_{nT}=i_T)P(S_{n1}=i_1,S_{n2}=i_2,\cdots,S_{nT}=i_T)\\&=\sum_{i_1=1}^{m}\cdots\sum_{i_T=1}^{m}\left[\prod_{t=1}^{T}P(Y_{nt}=y_{nt}|S_{nt}=i_t)\prod_{t=2}^{T}P(S_{nt}=i_t|S_{nt-1}=i_{t-1})\times\pi_n\right]\end{aligned} \tag{13}$$

A forward recursive algorithm can be applied by rearranging L_T in a more useful matrix form that follows MacDonald and Zucchini (1997):

$$L_{nT}=\pi_n\Lambda_n(i_1,y_1)Q_n(i_1,i_2)\Lambda_n(i_2,y_2)Q_n(i_2,i_3)\cdots Q_n(i_{T-1},i_T)\Lambda_n(i_T,y_T)1' \tag{14}$$

where n is the nth customer, $\Lambda_n(i_t,y_t)$ is an $m\times m$ diagonal matrix with $(P(y_{nt}|i_{nt}=1),\cdots,P(y_{nt}|i_{nt}=m))$ on the diagonal, π_n is a $1\times m$ vector of initial probability for each state, and $1'$ is an $m\times 1$ vector of ones. Therefore, the log-likelihood function for the HMM becomes the sum of individual log-likelihood, which can be represented as $\sum_{n=1}^{N}\ln L_{nT}$ in one-segment specification.

Multi-segment HMM Specification

In the multi-segment HMM, we overlay on top of our HMM a latent class specification across customers. This allows us to account for time-invariant customer heterogeneity that impacts behaviors but does not change with the customer–firm relationship (e.g. inherent preference for a particular channel, ability to pay). This is the model that we use in the upcoming empirical application, where we test a combination of static segments and dynamic states and select the model with the best fit and predication. The 1-segment HMM is a special case of the multi-segment HMM.

Conditional on the customer being in segment l and the unobservable learning process S_t, the observable process Y_t given l and S_t is independent. That is, for a group of T observations from a specific customer, the joint probability of an observed sequence of choices is now

$$\begin{aligned}&P(Y_1=y_1,Y_2=y_2,\cdots,Y_T=y_T)\\&=\sum_{\{l\}}\sum_{\{S\}}P(Y_1=y_1,Y_2=y_2,\cdots,Y_T=y_T|S_1=i_1,S_2=i_2,\cdots,S_T=i_T,l)\\&\quad\times P(S_1=i_1,S_2=i_2,\cdots,S_T=i_T|l)P(l)\\&=\sum_{l=1}^{L}P(l)\sum_{i_1=1}^{m}\cdots\sum_{i_T=1}^{m}\left[\prod_{t=1}^{T}P(Y_t=y_t|S_t=i_t,l)\prod_{t=2}^{T}P(S_t=i_t|S_{t-1}=i_{t-1},l)\right.\\&\quad\left.\times P(S_1=i_1|l)\right]=\sum_{l=1}^{L}ss_l L_{Tl}\end{aligned} \tag{15}$$

where $P(l)=ss_l$ is the proportion of segment l.

Therefore, the log-likelihood function for the multiple-segment HMM becomes

$$\ln L_{msHMM} = \sum_{n=1}^{N} \ln\left[\sum_{l=1}^{L} ss_l f_l(Y_{n1}, \cdots, Y_{nT})\right] = \sum_{n=1}^{N} \ln\left[\sum_{l=1}^{L} ss_l L_{nTl}\right] \tag{16}$$

for $n = 1, \cdots, N$, and $l = 1, \cdots, L$

where

$$ss_l = \frac{\exp(\pi_l)}{1+\sum_{l=1}^{L-1}\exp(\pi_l)} \tag{17}$$

$$L_{nTl} = \sum_{i_1=1}^{m}\cdots\sum_{i_T=1}^{m}\left[\prod_{t=1}^{T} P_n(Y_t = y_t|S_t = i_t, l)\prod_{t=2}^{T} P_n(S_t = i_t|S_{t-1} = i_{t-1}, l) \times P_n(S_1 = i_1|l)\right] \tag{18}$$

where n is the nth customer, l is the lth segment, π_1 is the size parameter, and ss_l is the likelihood of a consumer in segment l, which is the relative size of the lth segment ($0 < ss_l < 1, \sum ss_l = 1$). The parameters in L_{nTl} vary by segment.

Empirical Application

We apply our multivariate HMM to a longitudinal dataset of customers' observed purchases and channel choice from a multichannel clothing retailer. The dataset includes customer purchase behaviors as well as customer exposure to marketing communications. As it is typical in many CRM situations, this is a non-contractual setting in which relationship termination is not explicitly recorded. We first describe the data and the institutional details and then present the rationale for our choice of variables.

Data and Institutional Details

The dataset for this study is provided by a multichannel clothing retailer that sells its products through its network of brick-and-mortar stores, a catalog channel, and an online channel. Customer transaction information from multiple channels is captured and integrated. Thus, the company's CRM system can produce a complete multichannel purchase history for a particular customer. The dataset includes complete purchase history with marketing communication records for a sample of customers from December 2002 to July 2005. The firm in our study routinely conducts marketing communications through direct mailings to every member by sending new product information, flyers, promotions, and event notices.

We use the observations of the first 25 months to calibrate the model and the observations of the last 7 months for validation. For each customer, we have information on purchase amount, purchase time, channel choice (whether the purchase is made through retail store, online or catalog), and the number of marketing communication reminder mailings sent to that customer. In order to model relationship dynamics and retention rate, we choose customers who have made at least one purchase during the calibration period of December 2002 to December 2004, resulting in 595 customers in our calibration and validation datasets. Among 595 customers, there are 14,875 observations in the calibration period and 4165 observations in the validation period.

There are no price differences between the channels; in other words, there are no channel-specific price promotions. This practice of maintaining pricing consistency across channels is adopted by many successful retailers such as Nordstrom and the newly opened Amazon retail stores, ensuring that customers can choose different channels because of the service differentials of the channels instead of price. Further discussions with the management reveal that, during the time frame of the analysis, there is no targeting of marketing communication for the customers and that the mailings are not designed to drive customers to a specific channel. In order to model purchase incidence and to track retention, we divide the data into monthly increments to ensure a good data balance of 1s (purchases) and 0s (no purchase) in order to identify sufficient dynamics for managers to act upon. This practice is consistent among extant research that models purchase incidence (e.g. Netzer, Lattin, and Srinivasan 2008).[3]

Table 2 lists the descriptive statistics for our dataset. Among the 595 customers, each customer has an average of 8 purchases in the calibration period, ranging from 1 to 23. For each customer, the average dollar amount per purchase is \$40.02, ranging from \$11.5 to \$185, and the total purchase per person throughout the estimation is \$340.59. Out of all purchases, 84.01%, 13.53%, and 2.46% belong to offline retail, online, and catalog channels, respectively. The small number of online purchases is due to the fact that the online channel for the retailer was relatively new during the data period. Furthermore, customers receive an average of 2.9 marketing materials per month, and the proportion of customer making a purchase is 34.42% in the calibration sample and 25.57% in the hold-out sample.[4]

Variables

The variables for state-dependent choice behaviors include Z_{nt}, X_{nt}, and W_{nt}, which are assumed to influence the state-dependent purchase incidence, channel choice probability, and purchase amounts, respectively. These variables are distinguished from the variables constituting A_{nt} for the transition relationships. The vector A_{nt} is the set of variables that impact the transition probabilities and are assumed to have an enduring impact on customer retention and the customer–firm relationship, whereas

[3] Furthermore, in our dataset, we have checked a priori that there are no multiple purchases within the same month (thus no multiple channels in the same month).

[4] After later recovering the relationship states empirically, we find that, by the end of the calibration, 61.18% of the customers are in the inactive state and 38.82% of the customers are in the active state. This confirms the data description and leads us to believe that the company currently isn't doing a good job of keeping the customer active.

Table 2
Descriptive statistics (per person).

	Calibration sample	Holdout sample
Time frame	12/2002–12/2004	01/2005–07/2005
Number of months	25 months	7 months
Number of observations	14,875	4165
Number of customers	595	595
Average purchase incidences	8.61	1.8
Average amount per purchase	\$40.02	\$37.77
Total purchase amount per person	\$340.59	\$66.49
Number of marketing per month	2.9	2.9
If purchase is made, shop via		
Retail store	84.01%	80.47%
Online	13.53%	17.65%
Catalog	2.46%	1.90%

vectors Z_{nt}, X_{nt}, and W_{nt} are assumed to affect immediate behaviors.

Variables Affecting the Transition Matrix

Prior research has asserted that multichannel experiences impact customer relationships with firms (Kumar and Venkatesan 2005; Stone, Hobbs, and Khaleeli 2002). Previous research has also found that customer experience becomes more important as a relationship persists (Verhoef and Donkers 2005). Thus, channel-related experiences may impact customer–firm relationships. Also, recent research has demonstrated that it is possible to right-channel customer marketing messages (Montaguti, Neslin, and Valentini 2016), so channel experience is an item that can be influenced by marketing. In this study, we assume that relationship states are driven by customers' channel-related experiences and exposure to marketing communication, which would be entered as explanatory variables in the transition matrix.

We normalize channel experiences by dividing them by the time since the first purchase. This specification takes care of the variables' stationary property, accounts for cumulative impact, and incorporates forgetting while accounting for cumulative impact — the longer the customer has been with the firm without any activity, the smaller the cumulative experience becomes.[5]

We specify the relationship state as a function of normalized marketing communications and normalized cumulative purchases associated with each channel and define A_{nt} as follows:

$Retail\_exp_{nt}$ = normalized cumulative retail store purchases made by customer n by time $t-1$
$Online\_exp_{nt}$ = normalized cumulative online purchases made by customer n by time $t-1$
$Catalog\_exp_{nt}$ = normalized cumulative catalog purchases made by customer n by time $t-1$
$Marketing_{nt-1}$ = number of marketing communications sent to customer n at time $t-1$.

[5] We have also tried other normalization criteria, such as exponential decay on the cumulative experience, and find the proposed normalization method results in the best fit after accounting for model complexity.

Variables Affecting Purchase Incidence, Channel Choice, and Purchase Amount

Vectors Z_{nt}, X_{nt}, and W_{nt} are covariates that we believe to have immediate effects on state-dependent purchase incidence, channel choice, and purchase amounts, respectively. We allow the purchase amount to be influenced by channel experiences up to time $t-1$ and marketing at time t. We thus define Z_{nt}, X_{nt}, and W_{nt} as follows:

$Retail\_exp_{nt}$ = normalized cumulative retail store purchases made by customer n by time $t-1$
$Online\_exp_{nt}$ = normalized cumulative online purchases made by customer n by time $t-1$
$Catalog\_exp_{nt}$ = normalized cumulative catalog purchases made by customer n by time $t-1$
$Marketing_{nt}$ = number of marketing communications sent to customer n at time t.

Estimation Procedure and Model Selection

The parameters for our multi-segment, multivariate HMM are estimated by maximum likelihood estimation (MLE), which is accomplished through numerical optimization in GAUSS.

To justify the use of HMM and the associated additional sophistication to model dynamics, we test our specification against several benchmark models, namely,

a) A multinomial logit model without dynamic customer relationships or heterogeneity.
b) A 2-segment latent class model (LCA) without dynamic customer relationships.
c) A RFM model commonly used in B2C settings, where we use recency (inter-purchase time), frequency (the customer's cumulative number of purchases up to time t), and monetary value (the customer's historical average purchase amount to the current purchase) as independent variables to model purchase incidence, channel choice, and purchase amount.
d) A model based on observed state variables such as last channel choice (instead of channel experience) and marketing.
e) Our proposed multi-segment HMM model with different numbers of latent segments and latent states. The purpose of such an exhaustive search is to identify the best model fit.

In Table 3, we compare performance using the holdout sample log-likelihood in addition to BIC, which compares models and penalizes for model complexity. Based on these measures, the one-segment two-state HMM is the best-fitting model among all of the multi-segment, multivariate HMMs, and it also outperforms other commonly used benchmark models.

The result of the model comparison shows that 1) there are significant dynamics in buyer behavior, both in the short and long term, that cannot be captured by static between-customer heterogeneity; 2) the dynamics is gradual (and hence the customer preference evolution is gradual) and cannot be accurately captured by observed state variables such as previous channel choice and thus requires the need of a methodology such

Table 3
Selecting the number of segments and states & model comparison.

Model	Number of segments	Number of states	BIC	Holdout likelihood
Benchmark models				
(a) Logit w/o dynamics			25,462.54	−3284.85
(b) 2-segment LCA w/o dynamics			24,866.04	−3130.22
(c) RFM model			25,463.84	−3293.71
(d) Model w/ observed state variables			25,536.93	−3299.64
HMM	**1**	**2**	**24,657.51**	**−2987.43**
	1	3	24,693.75	−3010.37
	1	4	24,707.93	−3020.11
Multiple-segment HMM	2	2	24,876.21	−3017.54
	2	3	25,026.53	−3011.23
	3	2	25,008.94	−3014.65
	3	3	25,300.10	−3021.44
	4	2	25,168.67	−3022.58

The boldfaced data represents the data of proposed model.

as HMM to capture this latent dynamics; and 3) although the RFM model is often useful in predicting immediate decisions, it cannot inform us regarding the long-term impact of channel experience and marketing on customer behavior, as the model ignores their influence on the customer state transition.

Estimation Results

Table 4 shows the estimated parameters and corresponding standard errors for the one-segment two-state multivariate HMM. Table 5 highlights the characteristics of the two relationship states. Table 6 translates transition parameters into the transition matrix, with the covariates set at 0, showing the customer's intrinsic propensity to transition between the two states, and Table 7 shows the transition matrix with the covariates set at the mean levels. We first describe the states, followed by the discussion on the transition matrix.

The Two Relationship States

The relationship states which reflect different degrees of activity can be interpreted by examining the state-specific propensity to purchase (Table 4a). We calculate the intrinsic propensity to purchase by plugging the estimates in Table 4a into Eqs. (5) and (6) at the mean of the covariates: the purchase probability in state 1 and state 2 is respectively 21.75% and 77%. Table 5 describes the two states by averaging the purchase probability, the probability of making a purchase at store, online and catalog, and average purchase amount upon purchase. The table shows that, compared to customers in state 2, those in state 1 exhibit much lower purchase probability, are more likely to buy from a single channel (i.e. from retail store), and spend much less. The state 1 customer's preference for a store could suggest that these customers are unfamiliar with the retailer, and the retail store would be a safer place to buy (i.e. unfamiliarity might lead to buying the wrong product online), whereas the state 2 customer's preference for multichannel shopping indicates established trust and familiarity. Overall, this multidimensional view of the two relationship states implies that buyers in state 2 exhibit a stronger relationship with the retailer through more frequent purchases and higher purchase amounts relative to that of state 1. We therefore call state 1 the "inactive" state and state 2 the "active" state.

Table 4
Parameter estimates.

a — decision to purchase and channel choice

Parameter	State 1 Mean	State 1 Std Dev.	State 2 Mean	State 2 Std Dev.
Purchase probability				
δ	−2.3048	(0.068)	−1.9876	(0.083)
CV	0.1262	(0.013)	0.4087	(0.013)
$Retail\_exp_{nt}$	2.1215	(0.127)	5.5751	(0.395)
$Online\_exp_{nt}$	0.3487	(0.248)	3.7560	(0.543)
$Catalog\_exp_{nt}$	−2.9992	(0.308)	−0.0725	(0.099)
$Marketing_{nt}$	−0.2468	(0.043)	0.0222	(0.014)
Channel utility				
Retail	1.5308	(0.109)	2.1722	(0.240)
Online	−0.8459	(0.052)	2.1416	(0.242)
$Marketing_{nt\_retail}$	2.1006	(0.140)	0.0175	(0.055)
$Marketing_{nt\_online}$	−0.7556	(0.082)	0.0933	(0.049)
$Retail\_exp_{nt\_retail}$	1.8582	(0.067)	2.7594	(0.475)
$Retail\_exp_{nt\_online}$	−0.4947	(0.022)	2.3019	(0.459)
$Online\_exp_{nt\_retail}$	−0.5223	(0.103)	−0.0937	(0.084)
$Online\_exp_{nt\_online}$	0.2700	(0.009)	1.9459	(0.079)
$Catalog\_exp_{nt\_retail}$	−0.6499	(0.081)	−1.7515	(0.083)
$Catalog\_exp_{nt\_online}$	0.0213	(0.006)	−0.2634	(0.047)

b — purchase amount

Purchase amount Parameter	State 1 Mean	State 1 Std Dev.	State 2 Mean	State 2 Std Dev.
Intercept	3.0240	(0.074)	3.4528	(0.033)
$Retail\_exp_{nt}$	0.2986	(0.091)	0.0878	(0.075)
$Online\_exp_{nt}$	0.1269	(0.172)	0.0760	(0.072)
$Catalog\_exp_{nt}$	−0.2511	(0.101)	−0.2215	(0.045)
$Marketing_{nt}$	0.0099	(0.017)	0.0209	(0.007)

c — HMM transition parameter estimates

Transition matrix Parameter	State 1 Mean	State 1 Std Dev.	State 2 Mean	State 2 Std Dev.
μ	0.6717	(0.191)	0.0783	(0.020)
$Retail\_exp_{nt}$	0.4239	(0.040)	0.7223	(0.037)
$Online\_exp_{nt}$	0.2008	(0.197)	0.8166	(0.178)
$Catalog\_exp_{nt}$	−3.8153	(0.258)	−0.1716	(0.208)
$Marketing_{nt-1}$	0.0689	(0.020)	0.0261	(0.012)
Mean of initial probabilities	0.514		0.487	

Table 5
Description of the two HMM states.

	Inactive state	Active state
Purchase probability	21.75%	77%
If purchase is made, purchase via		
Store	97.00%	45.65%
Online	2.00%	52.05%
Catalog	1.00%	2.30%
Average purchase amount	$20.36	$44.53

Table 6
Transition matrix (intrinsic propensity to transition).

$t \rightarrow t+1$	Inactive	Active
Inactive	66.19%	33.81%
Active	51.96%	48.04%

The estimates of experience parameter (Table 4) indicate differences in the reaction to channel experiences across the two states. Although both retail and online channel experiences help increase the propensity to purchase for both relationship states, retail experience is the most influential one. Catalog experience has a negative impact on increasing purchase propensity. Marketing communications do not have an immediate impact on purchase incidence for both active and inactive customers.

We calculate the conditional probability of channel choice with all covariates set to mean values in order to examine the channel utility conditional on purchase across states. For a customer in an inactive state, the conditional probability of choosing a retail store is 97%, online is 2%, and catalog is 1% given purchase. The conditional probabilities given an active state are 45.65%, 52.05%, and 2.30% for retail, online, and catalog channels, respectively. A customer in an active state is more likely to make a purchase through multiple channels instead of a specific channel when a purchase occurs, whereas a customer in an inactive state is more likely to make a purchase at a retail store.

Inactive customers may not be familiar with the products and the company because they have lower purchase propensity, buy less frequently, and spend less. They may not remember the experience they had with the firm before. Therefore, they may prefer to buy in store to reduce the risk from product performance once they decide to make a purchase. Active customers demonstrate different channel tendencies. They are already familiar with the products and channels because they buy frequently. They don't receive much risk from product performance because of familiarity or brand loyalty. Therefore, buying through the online channel is more convenient for active customers who already have familiarity. Our result confirms that active customers show higher propensity to buy online, followed by purchasing in the retail store.

We next discuss transitions between the two states and the impact of marketing and channel experiences on migrations.

State Transitions

The parameter μ_j (Table 4c) represents the threshold between inactive state (state 1) and active state (state 2). The sign and absolute value of the threshold parameter imply how easily a customer moves from state 1 to state 2. The larger the value of the threshold, the less likely a jump from inactive to active is, and the more likely a jump from active to inactive is. A negative threshold value implies that it is easy to pass the threshold from inactive to active and thus that a customer is more likely to remain active when already in the active state or to switch toward activity when in the inactive state. A positive threshold value implies that a customer is more likely to switch toward inactivity or to remain inactive. A customer's intrinsic propensity to transition can be calculated by determining the threshold parameter μ_j with marketing and experience covariates set to zero (Table 6). In our application, the thresholds for each state are all positive, which represents the same information as intrinsic propensity to transition. When experience and marketing impacts are not taken into account, a customer in the inactive state has a higher intrinsic probability of remaining inactive (66.19%) than of migrating toward an active state (33.81%), whereas a customer in the active state has a higher intrinsic probability of moving toward an inactive state (51.96%) than of staying active (48.04%).

We then explore the impacts of the channel experience and marketing effects on the relationship (Table 7). The transition probabilities are calculated as the mean of the covariates. The propensity for entering an active state (41.19%) becomes higher than the intrinsic propensity to transition (33.81% — see Table 6) when a customer's current state is inactive, while the propensity for staying active increases to 56.30% from 48.04% when a customer is currently in an active state.

The coefficients for channel experiences and marketing communications represent the impact on transitions of, respectively, retail store, online, and catalog experiences and marketing communications received in the previous month. The sign of the coefficients implies whether channel experiences and marketing communications help a customer remain in or move toward an active state, and the value of the coefficients implies the magnitude of the influence. A negative coefficient means that the channel experiences or marketing communications increase the probability of being inactive, and a positive coefficient indicates migration towards the active state. We find that retail experience generally moves customers from inactive to active and also helps customers remain active. In the inactive state, retail experience is especially impactful for migrating customers from the inactive towards the active state, while the impact of online experience is limited. In the active state, retail and online experiences are equally effective in keeping active customers active. This confirms that the retail channel's rich experiential factors can help build familiarity (Verhoef, Neslin, and Vroomen 2007) for unfamiliar customers, whereas the online channel's limited effectiveness for inactive customers speaks to a lack of such rich interactions. The catalog, on the other hand, is detrimental in both immediate and long-term effects on customer transition to an active state. This speaks to the trend that the catalog is a sunset channel that no longer has a compelling value proposition — it has neither the retail channel's rich experience nor the online channel's convenience, and therefore the firm should phase it out.

The analysis in Table 8 illustrates the various impacts from channels and reinforces the statement above. Table 8a calculates

Table 7
Transition matrix (mean covariates).

$t \rightarrow t+1$	Inactive	Active
Inactive	58.81%	41.19%
Active	43.70%	56.30%

Table 8
Impact of channel experiences on retention probability.

a — the impact of channel experience on the probability from inactive to active
inactive (t − 1) → active (t)

Increase retail experience by	Probability	Increase online experience by	Probability	Increase catalog experience by	Probability
1	0.411	1	0.410	1	0.387
2	0.414	2	0.411	2	0.365
3	0.416	3	0.412	3	0.345
4	0.419	4	0.413	4	0.325
5	0.421	5	0.415	5	0.307
6	0.424	6	0.416	6	0.289
7	0.426	7	0.417	7	0.273
8	0.429	8	0.418	8	0.258

b — the impact of channel experience on the probability of remaining active
active (t − 1) → active (t)

Increase retail experience by	Probability	Increase online experience by	Probability	Increase catalog experience by	Probability
1	0.567	1	0.568	1	0.562
2	0.571	2	0.573	2	0.561
3	0.576	3	0.577	3	0.560
4	0.580	4	0.582	4	0.559
5	0.584	5	0.587	5	0.558
6	0.588	6	0.592	6	0.557
7	0.592	7	0.596	7	0.556
8	0.597	8	0.601	8	0.555

the probability from an inactive to an active state when the retail, online, and catalog experiences increase by 1 to 8 respectively, with other covariates remaining at the current value. We also normalized the channel experiences by tenure. More retail experiences help inactive customers migrate toward an active state, whereas more catalog experiences make customers become more inactive — the active probability changes from 0.387 to 0.258 when catalog experience increases the current level by 8. Table 8b shows the probability of retention (active to active) when three channel experiences increase from 1 to 8 separately. Increasing retail and online experiences can increase the retention probability from 0.57 to 0.60 when either retail or online experience increases the current level by 8. More catalog experience decreases the retention probability for active customers, but the magnitude of decrease is not as great as that of inactive customers.

We also calculate the long-run stationary probability for the transition matrix to show the impact of marketing by increasing the number of marketing communications from 1 to 8 (Table 9).

Table 9
Impact of marketing on retention.

Increase marketing communications by	Inactive prob.	Active prob.
1	51.67%	48.33%
2	50.30%	49.70%
3	48.96%	51.04%
4	47.65%	52.35%
5	46.37%	53.63%
6	45.11%	54.89%
7	43.88%	56.12%
8	42.69%	57.31%

The long-run active probability can be seen as retention probability. Increasing number of marketing campaigns will increase the retention probability and decrease the inactive probability. Marketing communications can play a role as a reminder or informational media which helps customers migrate away from an inactive state and help customers remain active.

Counterfactual Analysis

A customer's relationship state at any given period of time can be recovered probabilistically through "smoothing" or "filtering". The goal is to recover the relationship state at the ending period T and to use the full information available up to time T to recover the relationship state. It is helpful to know the state of a customer's relationship at the end of the observation period, because marketers can then build differentiated strategies for the future period based on that state. In this section, we recover customer relationship states at the end of the observation period (December 2004) and simulate a 7-month counterfactual plan for marketing communications. We then compare the performance between the base case and the counterfactual marketing plan.

The recovery of the relationship state at the end of an observation period is calculated by using Eq. (i) in Appendix A. In December 2004, 38.82% of customers were in the active state, and 61.18% were in the inactive state. We then use forward simulation to generate a sequence of purchase events in the 7-month horizon. Each purchase event in the simulated sequence is characterized by purchase decision, channel decision, and purchase amount decision as well as the updated state-membership probabilities. The updated state-membership probabilities will be used to generate the decision in the next purchase event.

We now compare the performance of the following two scenarios:

(1) Base case — This uses the company's current marketing efforts as seen in the data.
(2) Marketing optimization — In this scenario, the optimized number of marketing communications is determined for each customer based on her state membership.

In scenario (1), the firm keeps following the current marketing efforts and does not make any extra effort to improve the customer–firm relationship. The number of purchase incidence is 1065, and the total purchase amount is $39,560. The gross profit margin in the calibration data is 49%, and we apply this margin in both scenarios. The gross profit in scenario (1) is $19,384.40. The total number of direct mail marketing communications sent to 595 customers is 8812. The cost of a direct mail is assumed to be $1 per piece in our analysis,[6] which covers printing, art and preparation, consultation fees, computer processing, etc. The total costs for 8812 marketing communications would be $8812, and therefore the profit would be $10,572.40 (Table 10).

In scenario (2), we optimize marketing for each of the two states based on the parameter estimates of HMM in order to maintain the relationship and improve the customer–firm relationship in the 7-month simulation periods. Based on the customer's current state, the firm will increase the level of marketing in order to move them to the active state. Once the customer reaches the active state, then the marketing effort will be back to the baseline current level. Hence, the optimized marketing will be applied only to inactive customers. Table 10 shows that the optimized marketing policy outperforms the base case and generates 140% improvement in profits when compared to the base policy ($25,359 vs. $10,572.40). This tremendous improvement in profitability comes from the increased number of active customers as the result of marketing optimization, their characteristic high purchase probability, and the higher purchase amount.

Table 10
Policy simulation.

Policy	Base case	Marketing optimization
	1	2
Total purchase incidences	1065	2120
Total revenue	$39,560	$84,800
Gross profit margin	49%	49%
Gross profit	$19,384.40	$41,552.00
Marketing expense	$8812.00	$16,193.00
Profit	$10,572.40	$25,359.00

Conclusions and Directions for Future Research

We have proposed a framework to estimate multichannel customer retention in a non-contractual setting. The proposed multivariate HMM simultaneously models the changes in purchase incidence, channel choice, and purchase amount with respect to relationship states and investigates the impact of marketing communications on state transition. The model advances prior research on customer retention in non-contractual service settings by 1) modeling retention probabilities as driven by channel experiences and marketing, 2) modeling multiple decisions of high managerial interests, 3) incorporating channel preference evolution, and 4) flexibly allowing customer "revival" from inactive to active states.

In the empirical application, we use observed choice behavior and purchase information to make inferences about customers' underlying relationship states and identify two dynamic relationship states that differ in retention and profitability. Compared to the "inactive" state, customers in the "active" state not only purchase more frequently and are more likely to stay active (thus demonstrating higher retention), but also spend more per purchase, resulting in much higher profitability. Customers in the active states also tend to be multichannel customers, which is consistent with prior research, which states that multichannel customers tend to appreciate service offerings of various channels and hence are more profitable.

We find that customers' channel experiences have both short- and long-term impacts on relationship dynamics and state-dependent choice. More importantly, the influences of different channel experiences are asymmetric across relationship states. Table 4a, b and c shows that for customers in the inactive state, offline experience is far more impactful than online experience in increasing purchase incidence and purchase amount, and in migrating customers from the inactive state to the active state. Once the customers get to the active state, however, the influence of offline experience decreases, and the online experience becomes more impactful. These findings stem from the service output differentials offered by the various channels. Previous research suggests that retail stores provide rich sensory experiences that can produce a range of psychological and behavioral outcomes, can serve as powerful memory aids (Balasubramanian, Raghunathan, and Mahajan 2005; Bitner 1992; Raghunathan and Irwin 2001; Schmitt 2000), and the full service retailer can assist customers in every phase of the shopping processes to create unique store experiences (Kerin, Hartley, and Rudelius 2012). For new customers or existing customers who have lapsed into the inactive state, rich store experiences could stimulate them by providing either an educational environment to learn about product attribute or service outputs to make the retailer's value proposition salient, thereby building stronger customer–firm relationship. Compared to the offline stores, the online channel offers lower transactional cost in the form of convenience (Chintagunta, Chu, and Cebollada 2012). Active customers who already see the value of the retailer and therefore do not need the immersive store experience would have higher appreciation for the online channel.

[6] The $1 figure is conservatively set at the higher range of most direct marketing efforts, as we don't want to make direct marketing expense trivial.

In fact, we do find empirical evidence that the offline (retail-store) channel is more useful for migrating customers from the inactive state to the active state, effectively serving an “educational” or “revival” purpose. Similarly, the online channel is most effective in keeping existing active customers active, thus serving a “retention” purpose. The catalog, on the other hand, is inferior in both immediate and long-term effects on customer activity and profitability. This speaks to the trend that the catalog is a sunset channel that no longer has a compelling value proposition; it has neither the retail channel’s rich experience nor the online channel’s convenience. Our research provides evidence for the roles of channel experiences in managing the customer lifecycle. Using our modeling framework, firms can identify the latent customer relationship states, and then encourage certain channel participation to build long and profitable relationships.

We also find that marketing communication has limited impacts on immediate behaviors but is effective in migrating customers to the active state, suggesting a lag in customer’s processing of the reminder information. This finding also suggests that firms should adopt a longer term perspective when evaluating the marketing effectiveness. Using the counterfactual analysis shown in Tables 8, 9, and 10, we highlight the opportunity for the multichannel firm to use marketing optimization to dynamically manage and increase the retention and hence the value of its customer base. Our model allows multichannel marketers to more accurately predict customer retention in non-contractual settings and assess how different channels and marketing influence customer retention and profitability. Furthermore, firms can use the model to identify active and inactive customers in its portfolio and design differential marketing strategies for active customers vs. inactive customers, with the goal of migrating customers to and retaining customers in the active state.

Accurately classifying customers into both time-invariant segments and dynamic relationship state is important for a firm to optimize its targeted marketing efforts such as differentiated pricing, promotions, and communications strategies. Possible future applications for our framework include any multichannel retailer or service providers (e.g. such as banks) that seek to accurately estimate retention and wish to understand how the use of different marketing strategies impact the client–firm relationship evolution and retention.

We now list the limitations of this research and offer avenues for further research. First, as is typical in most CRM datasets, our study only considers channel experience from our focal seller and does not account for channel experiences from competitors. Future extension may investigate the impact of competitors’ activities on relationship dynamics. Another limitation is that the company in our data frame has conducted only direct mailing marketing communications. Future studies can assess the impact of various communication media on relationship evolution and choice preference. Email marketing communication may have different impacts on customer retention and profitability than direct mailings. Finally, the specific content of the firm’s direct mail marketing communications is not clear (i.e. we do not know whether the mailing notices are for annual sales or holiday sales, flyers for new product introduction, or coupons for membership reward), and our data does not reveal the firm’s goal for communication. Further research can address how various content and goals of marketing communications may differentially impact multi-channel customer retention.

Appendix A. Recovering State Membership

A customer’s relationship state at any given period of time can be recovered probabilistically through “smoothing” or “filtering”. Given a customer’s history of observed behavior from period 1 to T, the probability distribution of the relationship state for the nth customer at time T is as follows.

$$P(S_{nT}=i|Y_{n1},Y_{n2},\cdots,Y_{nT}) = \pi_n\Lambda_n(i_1,y_1)Q_n(i_1,i_2)\Lambda_n(i_2,y_2)Q_n(i_2,i_3)\ldots \tilde{Q}_n(i_{T-1},i_T=i)\tilde{\Lambda}_n(i_T=i,y_T)/L_{nT} \quad (i)$$

where $\Lambda_n(i_t,y_t)$ is an $m\times m$ diagonal matrix with $(P(y_{nt}|i_{nt}=1),\cdots,P(y_{nt}|i_{nt}=m))$ on the diagonal, π_n is a $1\times m$ vector of initial probability for each state, $\widetilde{Q}_n(i_{T-1},i_T=i)$ is the ith column of $Q_n(i_{T-1},i_T)$, $\tilde{\Lambda}_n(i_T=i,y_T)$ is $P(y_T|i_T=i)$,T is the end of the observation period, and L_{nT} is the joint probability of an observed sequence of choices Y per Eq. (13).

References

Balasubramanian, Sridhar, Rajagopal Raghunathan, and Vijay Mahajan (2005), “Consumers in a Multichannel Environment: Product Utility, Process Utility, and Channel Choice,” *Journal of Interactive Marketing*, 19, 2, 12–30.

Bitner, Mary Jo (1992), “Servicescapes: The Impact of Physical Surroundings on Customers and Employees,” *Journal of Marketing*, 56, 2, 57–71.

Blattberg, Robert C. and John Deighton (1996), “Manage Marketing by the Customer Equity Test,” *Harvard Business Review*, 74, 4, 136–44.

Boehm, Martin (2008), “Determining the Impact of Internet Channel Use on a Customer’s Lifetime,” *Journal of Interactive Marketing*, 22, 3, 2–22.

Campbell, Dennis and Frances Frei (2010), “Cost Structure, Customer Profitability, and Retention Implications of Self-service Distribution Channels: Evidence from Customer Behavior in an Online Banking Channel,” *Management Science*, 56, 1, 4–24.

Chintagunta, Pradeep K., Junhong Chu, and Javier Cebollada (2012), “Quantifying Transaction Costs in Online/Off-line Grocery Channel Choice,” *Marketing Science*, 31, 1, 96–114.

Du, Rex Y. and Wagner A. Kamakura (2006), “Household Life Cycles and Lifestyles in the United States,” *Journal of Marketing Research*, 43, 1, 121–32.

Fader, Peter S., Bruce G.S. Hardie, and Ka Lok Lee (2005), ““Counting Your Customers” the Easy Way: An Alternative to the Pareto/NBD Model,” *Marketing Science*, 24, 2, 275–84.

——— and ——— (2010), “Customer-base Valuation in a Contractual Setting: The Perils of Ignoring Heterogeneity,” *Marketing Science*, 29, 1, 85–93.

Greene, William (2003), *Econometric Analysis*. Upper Saddle River, New Jersey: Prentice Hall.

Gupta, Sunil and Donald R. Lehmann (2003), “Customers as Assets,” *Journal of Interactive Marketing*, 17, 1, 9–24.

Hansotia, Behram J. and Bradley Rukstales (2002), “Direct Marketing for Multichannel Retailers: Issues, Challenges and Solutions,” *Journal of Database Marketing*, 9, 3, 259–66.

Hitt, Lorin M. and Frances X. Frei (2002), “Do Better Customers Utilize Electronic Distribution Channels? The Case of PC Banking,” *Management Science*, 48, 6, 732–48.

Kerin, Roger, Steven Hartley, and William Rudelius (2012), *Marketing*. New York, NY: McGraw-Hill Education.

Kumar, V. and Rajkumar Venkatesan (2005), "Who Are the Multichannel Shoppers and How Do They Perform?: Correlates of Multichannel Shopping Behavior," *Journal of Interactive Marketing*, 19, 2, 44–62.

Lal, Rajiv and Miklos Sarvary (1999), "When and How is the Internet Likely to Decrease Price Competition?" *Marketing Science*, 18, 4, 485–503.

Li, Shibo, Baohong Sun, and Alan L. Montgomery (2011), "Cross-selling the Right Product to the Right Customer at the Right Time," *Journal of Marketing Research*, 48, 4, 683–700.

Luo, Anita and V. Kumar (2013), "Recovering Hidden Buyer–Seller Relationship States to Measure the Return on Marketing Investment in Business-to-Business Markets," *Journal of Marketing Research*, 50, 1, 143–60.

MacDonald, Iain and Walter Zucchini (1997), *Hidden Markov and Other Models for Discrete-valued Time Series*. London; New York: Chapman & Hall.

Montaguti, Elisa, Scott A. Neslin, and Sara Valentini (2016), "Can Marketing Campaigns Induce Mutichannel Buying and More Profitable Customers? A Field Experiment," *Marketing Science*, 35, 2, 201–17.

Montoya, R. Matthew, Robert S. Horton, and Jeffrey Kirchner (2008), "Is Actual Similarity Necessary for Attraction? A Meta-analysis of Actual and Perceived Similarity," *Journal of Social and Personal Relationships*, 25, 889–922.

Montoya, Ricardo, Oded Netzer, and Kamel Jedidi (2010), "Dynamic Allocation of Pharmaceutical Detailing and Sampling for Long-term Profitability," *Marketing Science*, 29, 5, 909–24.

Netzer, Oded, James M. Lattin, and V. Srinivasan (2008), "A Hidden Markov Model of Customer Relationship Dynamics," *Marketing Science*, 27, 2, 185–204.

Pfeifer, Phillip E. and Paul W. Farris (2004), "The Elasticity of Customer Value to Retention: The Duration of a Customer Relationship," *Journal of Interactive Marketing*, 18, 2, 20–31.

Raghunathan, Rajagopal and Julie R. Irwin (2001), "Walking the Hedonic Product Treadmill: Default Contrast and Mood-based Assimilation in Judgements of Predicted Happiness with a Target Product," *Journal of Consumer Research*, 28, 3, 355–68.

Rangaswamy, Arvind and Gerrit H. Van Bruggen (2005), "Opportunities and Challenges in Multichannel Marketing: An Introduction to the Special Issue," *Journal of Interactive Marketing*, 19, 2, 5–11.

Reichheld, Frederick F. and W. Earl Sasser Jr. (1990), "Zero Defections: Quality Comes to Services," *Harvard Business Review*, 68, 5, 105–11.

Reinartz, Werner J. and V. Kumar (2000), "On the Profitability of Long-life Customers in a Noncontractual Setting: An Empirical Investigation and Implications for Marketing," *Journal of Marketing*, 64, 4, 17–35.

Schmitt, Bernd H. (2000), *Experiential Marketing*. New York: Free Press.

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

Schweidel, David A., Peter S. Fader, and Eric T. Bradlow (2008), "A Bivariate Timing Model of Customer Acquisition and Retention," *Marketing Science*, 27, 5, 829–43.

———, Eric T. Bradlow, and Peter S. Fader (2011), "Portfolio Dynamics for Customers of a Multiservice Provider," *Management Science*, 57, 3, 471–86.

Stone, Merlin, Matt Hobbs, and Mahnaz Khaleeli (2002), "Multichannel Customer Management: The Benefits and Challenges," *Journal of Database Marketing*, 10, 1, 39–52.

Verhoef, Peter C. and Bas Donkers (2005), "The Effect of Acquisition Channels on Customer Loyalty and Cross-buying," *Journal of Interactive Marketing*, 19, 2, 31–43.

———, Scott A. Neslin, and Björn Vroomen (2007), "Multichannel Customer Management: Understanding the Research-shopper Phenomenon," *International Journal of Research in Marketing*, 24, 2, 129–48.

Zhang, Jonathan Z., Oded Netzer, and Asim Ansari (2014), "Dynamic Targeted Pricing in B2B Relationships," *Marketing Science*, 33, 3, 317–37.