Some Customers Would Rather Leave Without Saying Goodbye

Eva Ascarza
Oded Netzer
Bruce G. S. Hardie†

April 2017

†Eva Ascarza is Associate Professor of Marketing, Columbia Business School (email: ascarza@gsb.columbia.edu). Oded Netzer is Associate Professor of Marketing, Columbia Business School (email: onetzer@gsb.columbia.edu). Bruce G. S. Hardie is Professor of Marketing, London Business School (email: bhardie@london.edu). The authors benefited from comments by Kamel Jedidi, the audience at Temple University, the brown bag seminar series at the Columbia Business School, and the participants of the 2013 Marketing Science conference, the 2013 Joint Statistical Meetings, and the 2016 European Marketing Academy conference.

Abstract

Some Customers Would Rather Leave Without Saying Goodbye

We investigate the increasingly common business setting in which companies face the possibility of both observed and unobserved customer attrition (i.e., "overt" and "silent" churn) in the same pool of customers. This is the case for many online-based services where customers have the choice to stop interacting with the firm either by formally terminating the relationship (e.g., cancelling their account) or by simply ignoring all communications coming from the firm. The standard contractual versus noncontractual categorization of customer-firm relationships does not apply in such hybrid settings, which means the standard models for analyzing customer attrition do not apply. We propose a hidden Markov model (HMM)-based framework to capture silent and overt churn. We apply our modeling framework to two different contexts — a daily deals website and a performing arts organization. In contrast to previous studies that have not separated the two types of churn, we find that overt churners in these hybrid settings tend to interact more, rather than less, with the firm prior to churning. That is, in settings where both types of churn are present, a high level of activity — such as customers actively opening emails received from the firm — is not necessarily a good indicator of future engagement; rather it is associated with higher risk of overt churn. We also identify a large number of "silent churners" in both empirical applications — customers who disengage with the company very early on, rarely exhibit any type of activity, and almost never churn overtly. Furthermore, we show how the two types of churners respond very differently to the firm's communications, implying that a common retention strategy for proactive churn management is not appropriate in these hybrid settings.

Keywords: Churn, retention, attrition, customer relationship management, customer base analysis, hidden Markov models, latent variable models.

1 Introduction

Understanding and managing customer retention is one of the most important aspects of customer relationship management (Blattberg et al. 2008). As companies become more proactive in their management of customer retention, managers are increasingly interested in identifying those customers at risk of leaving the firm (i.e., churning) and in finding ways to reduce such churn. It is common to make the distinction between identifying and predicting churn in contractual versus noncontractual settings. In contractual settings, the existence of a contract between the customer and the firm means the firm will know when the customer has ended her relationship with the firm. In contrast, customers in noncontractual settings do not explicitly signal to the firm that they wish to terminate their relationship; as a result, the firm does not observe churn, only an absence of behavior.

In the digital era, with the proliferation of free and freemium type pricing schemes, customers often have a formal relationship with firm (e.g., an account or subscription) but do not need to continue paying the firm to keep the relationship alive. Examples include services such as daily deals sites (e.g., Groupon), social networks (e.g., LinkedIn), online games (e.g., Candy Crush), or web-based email services (e.g., Gmail). In such settings, customers can choose whether to churn overtly by informing the company that they wish to terminate their relationship with the firm (e.g., closing the account) or just silently disappear without saying goodbye (e.g., never interacting with the firm again). These settings constitute a hybrid of the standard contractual and noncontractual categorization of customer-firm relationships. Treating such hybrid settings as noncontractual, thus ignoring overt churn (e.g., account closure), misses the opportunity to identify, predict, and manage overt churners. On the other hand, treating such hybrid setting as contractual, thus ignoring latent attrition, erroneously assumes that all customers who did not inform the firm about their silent termination of the relationship are indeed active customers.

In this paper we explore such hybrid settings by analyzing customer behavior in two different contexts: a daily deals website and a performing arts organization. By separating silent churners from overt churners we are able to not only understand and predict both types of churn, but also explore possible levers to better manage the customer base. In contrast to previous studies, which have not separated the two types of churn, we consistently find that overt churners tend to interact more, rather than less, with the firm. In other words, in hybrid settings, a high level of activity (represented in our empirical applications by frequently opening the emails received from the firm) may not necessarily imply high levels of engagement in the future. Rather, it might be associated with a higher risk of overtly churning. Despite the differences across the two contexts investigated in this research, we consistently find that, while overt churners tend to frequently consider the firm's communications (i.e., open emails received from the firm), they rarely engage with the content (e.g., they do not click on links in the emails). This suggests that such customers are not finding the right offering(s) and therefore, after repeated disappointments, terminate their relationship with the firm. We also find a large number of silent churners in both empirical applications. These are customers who disengage from the company very early on, rarely open emails and almost never explicitly unsubscribe from the service. In other words, many customers are leaving without saying goodbye.

In order to accommodate and separate both types of churn we develop a hidden Markov model (HMM)-based framework that allows us to capture different latent states of customer behavior: those customers who are at risk of overtly churning (captured by an "at risk" state), those who have silently churned (captured by a "silently gone" state), and those who plan to continue interacting with the firm (captured by an "engaged" state). The model allows customers to transition among these states, thereby capturing dynamics in their relationship with the firm. Furthermore, we allow factors such as the content and quality of the firm's communications to differentially affect customers' transitions among (latent) states, as well as impact their behavior (given state membership).

Using our model, we demonstrate that different firm actions are needed in order to mitigate the risk of overt and silent churn. Specifically, we find that sending communications with "better" content to customers in the "at risk" (of overtly churning) state can encourage them to move to a state of engagement where they are more likely to click on the email and less likely to overtly churn. Moreover, sending the "better" content (e.g., content customized to better match customers' preferences) reduces the probability that engaged customers will silently leave. Managing those who have already left silently is more challenging. We find that once a customer transitions to the "silently gone" state, the firm is highly unlikely to re-engage the customer using previously used communication methods. These findings have important implications for marketers. Firms should either find creative ways to re-engage these customers or simply let them go (i.e., stop sending them communications).

We continue by discussing the relevant literature on overt and silent churn (Section 2). In Section 3, we develop a model that accommodates both types of churn in hybrid settings. In Sections 4 and 5 we apply our model to two different contexts, a daily deals website and a performing arts organization, and show how the model clearly separates the two types of churn. Using counterfactual analyses, we also demonstrate how the firm could differentially mitigate the two types of churn by changing the quality of its offerings and the type of communications with its customers. We conclude (Section 6) with a discussion of the theoretical and practical contributions of this paper and directions for future research.

2 Overt Versus Silent Churn

Following Schmittlein et al. (1987), it is common to characterize the nature of a firm's relationship with its customer base in terms of whether or not the loss of a customer is observed by the firm. The term *contractual* is used for those settings in which the loss of the customer is observed by the firm. These include both settings where the "default" is that the provision of the service will stop at a specific point in time (known in advance to firm and the customer) unless the customer takes a specific action (e.g., renews a subscription or membership), and those settings where the "default" is that the firm continues to provide the service until the customer contacts the firm to cancel her contract. The term *noncontractual* is used for those settings where the loss of the customer is not observed by the firm (Reinartz 1999). There is no explicit signal from the customer to the firm that she has stopped being a customer; instead, she "silently attrites" (Mason 2003).

The rich literature on customer attrition has modeled churn according to this dichotomized categorization. In contractual settings, (observed) attrition is often modeled in terms of whether or not the customer churns at the end of each period using logistic regression or more general data-mining techniques (e.g., Ascarza 2016, Coussement and Van den Poel 2008, Lemmens and Croux 2006, Neslin et al. 2006) or in terms of how long she remains a customer using a duration-time model (e.g., Fader and Hardie 2007, Schweidel et al. 2008). The majority of this literature has focused on identifying predictors of churn, which typically consist of customer characteristics (e.g., demographics), past interactions with the firm (e.g., complaints, product returns, calls to the customer service), and measures of past usage, generally summarized in terms of the recency and frequency of past activity. While the exact operationalization of the activity variables and their effect sizes vary among industries and across studies, a consistent finding in contractual settings is that a lack of (or decline in) customer activity is a strong predictor of customer loss (e.g., Ascarza and Hardie 2013, Borle et al. 2008, Buckinx et al. 2007, Chen et al. 2015, Coussement and De Bock 2013, Lemmens and Croux 2006).

In noncontractual settings, attrition is by definition unobserved. As such, it is not modeled directly but must be inferred from a lack of observed transaction activity. Models such as the Pareto/NBD (Schmittlein et al. 1987), the BG/NBD (Fader et al. 2005) and the BG/BB (Fader et al. 2010) assume customers behave as if they transact randomly around their mean propensity until they "die" at some time that is unobserved by the firm. The resulting model can then be used to infer the probability that a customer with a given transaction history is still "alive." Because identifying churn is already a challenge (as it is unobserved), the literature on noncontractual settings has mostly focused on identifying *which customers* are most likely to have "died" (i.e., silently churned), rather than identifying *which factors* precede such latent attrition. Notable exceptions are Braun et al. (2015), Knox and Van Oest (2014), Schweidel and Knox (2013), and Schweidel et al. (2014), which build on the aforementioned latent attrition models and allow for covariates such as direct marketing activity and characteristics of customer transactions to impact the probability of (silently) churning.

While many businesses fit a binary contractual/noncontractual categorization of customer-firm relationships, there are many hybrid settings where the loss of some customers is observed by the firm while the loss of others is unobserved. As such, it is the customer who decides whether to churn latently or overtly. While such phenomena is not entirely new,[1] the coexistence of observed and unobserved attrition has not been explored. Developments in information technology and the proliferation of digital services imply that the number of such hybrid settings is rapidly growing. For example, in the days of traditional catalog retailing, the cost to customers of telling the retailer that they were no longer interested in receiving the catalogs was sufficiently high that they "silently attrited." Nowadays, customers can simply click on the unsubscribe link at the bottom of any email communications received from the retailer to "formally" churn. Many online services (e.g., Facebook, LinkedIn, Ebay) as well as most mobile games (e.g., CandyCrush) belong to the category of hybrid settings. For example, a Facebook user who no longer wishes to use the service can either explicitly close her account or simply stop logging on.

[1] Schmittlein et al. (1987) note that bank accounts, which may typically be thought of as being contractual in nature, can sometimes become permanently dormant due to customers forgetting or even physically dying, and the bank being unaware of the customer churn. No-fee credit cards may be another such example.

Despite the prevalence of such hybrid settings in today's economy, to the best of our knowledge, the duality between observed and unobserved attrition has not been investigated.[2] Accordingly the objectives of this research are to highlight the coexistence of both types of churn in many business settings, to investigate the dynamics between customer activity and each of these churn types, and to explore how firms can manage churn (both overt and silent) in such hybrid settings. To do so, we build a model that not only accommodates but also separates both types of churn (based on other behaviors observed by the firm) and allows factors such as marketing actions to affect customer behaviors in the short-run, as well as the medium-to-long term.

3 The Model

We generalize the existing methods for modeling customer attrition by proposing a modeling framework that can accommodate both observed and unobserved churn. In particular, we use an HMM as a unifying framework suitable for both types of churn.

There is a long tradition of using (manifest) Markov models to characterize buyer behavior (Fader et al. 2014), and they have been used to model the behavior of a customer base. For example, Deming and Glasser (1968) model the duration of customer subscriptions in a contractual setting using a Markov chain. Dwyer's (1989) customer migration model of behavior in noncontractual settings can be viewed as a Markov model in which the observed states are characterized by the customer's past activity summarized in terms of recency, frequency, and monetary value (RFM) (Pfeifer and Carraway 2000). More recently, a number of researchers (e.g., Ascarza and Hardie 2013, Netzer et al. 2008, Schweidel et al. 2014, Schweidel et al. 2011) have used hidden/latent Markov models to capture dynamics in the nature of the relationship between customers and the firm in both contractual and noncontractual settings.

As mentioned above, the standard latent attrition models for noncontractual settings (e.g., Schmittlein et al. 1987; Fader et al. 2010) capture dynamics in behavior via a latent absorbing state: customers transact/spend at a constant rate until they become permanently inactive (at which time their transaction/spend rate goes to zero). While they have been extended to capture multiple behaviors (e.g, Schweidel et al. 2014), they do not allow for dynamics in behavior while

[2]Note that the duality investigated in this paper is different from that discussed in Braun and Schweidel (2011): voluntary and involuntary churn. In their case, both types of churn are observed by the firm and the difference lies on *who initiates* the termination of the relationship (i.e., the customer or the firm). In hybrid settings attrition is always the customer's choice; the difference lies on whether or not the company observes when the customer churns.

the customer is "alive." So rather than starting with a latent attrition model and adding on churn as a possible behavior while the customer is "alive," we take the more flexible approach of building our model using an HMM.[3] This allows for a richer set of dynamics in customer behavior. For example, customers can change their propensity to unsubscribe or their propensity to engage with the service.

3.1 Underlying Logic

We seek to model the multidimensional behavior of customers over time. Each period we observe whether or not the customer has terminated their relationship with the firm (unsubscribing, in the examples used in our empirical applications) as well as their behavior on one or more dimensions of interest (opening emails and clicking on links, in the examples used in our empirical applications). In the case of an app, app deletion would be the signal of relationship termination and the other behaviors of interest could include time spent using the app and in-app purchases. In a social network context, termination would be captured by account closure, and activities could include connection requests, posts, likes, and so on. Because we observe the same behaviors in both of the empirical applications presented in Sections 4 and 5 — open, click, and unsubscribe — we will, for expositional clarity, develop our model in terms of these three behaviors. (The model can easily be modified to accommodate other behaviors.)

More formally, we have a trivariate binary random vector $\boldsymbol{Y}_{it} = [\,Y^o_{it}, Y^c_{it}, Y^u_{it}\,]$ with realization $\boldsymbol{y}_{it} = [\,y^o_{it}, y^c_{it}, y^u_{it}\,]$, where $y^o_{it} = 1$ if individual i *opens* the email sent in period t (0 otherwise), $y^c_{it} = 1$ if individual i *clicks* on any link in the email sent in period t (0 otherwise), and $y^u_{it} = 1$ if individual i *unsubscribes* via the email opened in period t (0 otherwise).

We assume the existence of a set of latent states that reflect the nature of an individual's relationship with the firm, and model $P(\boldsymbol{Y}_{it} = \boldsymbol{y}_{it})$ as a function (in part) of the latent state occupied by individual i in period t. For example, we would expect that a customer who is at risk of terminating her relationship with the firm will be captured by a state that exhibits a high probability of unsubscribing. In contrast, a customer who is engaged with the firm and has a low risk of churning will be captured by a state characterized by high levels of activity (opening and clicking) and a low probability of unsubscribing. Finally, we would expect that a "silent churner"

[3] Note that latent attrition models can be formulated as (constrained versions of) a two-state HMM where one of the states is an absorbing state of (latent) attrition (Schwartz et al. 2014).

will be captured by a state characterized by both very low levels of activity (opening and clicking) and a low probability of unsubscribing.

It is important to note that while we could impose additional structure on the latent states — e.g., latent attrition being captured by an absorbing state that allows no activity (e.g., Schwartz et al. 2014) — we instead allow the data to inform us about what types of relationship states exist and the nature of customer behavior given the state.[4] We assume that individuals transition among the latent states following a first-order Markov process, thus capturing when customers transition from an "active" to an "inactive" state, or to a state with a high risk of termination. Furthermore, to better understand the antecedents and consequences of the two types of churn, we allow managerially relevant covariates such as the quality of the firm's offerings and type of communications with customers to affect customers' transitions among states as well as their behaviors given membership of a particular state.

3.2 Model Specification

The model comprises two main components, both occurring at the individual level: (i) the "relationship state" evolution, and (ii) the customer's state-dependent behaviors (e.g., the probability of opening, clicking, and unsubscribing). To avoid spurious inferences about dynamics in customer behavior, we account for heterogeneity in the parameters of both the state evolution and the customer behavior (given state membership) processes.

3.2.1 Relationship State Evolution

We assume K hidden/latent/unobserved relationship states, which differ with respect to the level of expected activity and the risk of the customer terminating their relationship with the firm.[5] In order to capture nonstationarity in underlying customer behavior, we allow customers to transition among the latent states over time. In its most general form, we assume that S_{it}, the state occupied by customer i in period t, evolves over time following a (hidden) Markov process with heterogeneous

[4]We estimated an "absorbing" version of our main model in which one of the states is forced to be absorbing and all three behaviors in this state are forced to zero (no activity). The fit of this model specification was inferior to that of our proposed model specification.

[5]Consistent with prior research using HMMs (e.g., Netzer et al. 2008, Schweidel et al. 2011), we assume that the number of latent states is common across all customers.

nonstationary transition matrix $\boldsymbol{Q}_{it}$, defined as

$$\boldsymbol{Q}_{it} = \begin{pmatrix} q_{i11t} & \cdots & q_{i1Kt} \\ \vdots & \ddots & \vdots \\ q_{iK1t} & \cdots & q_{iKKt} \end{pmatrix},$$

where

$$P(S_{it} = k' | S_{i(t-1)} = k) = q_{ikk't}\,, \;\; k, k' \in \{1, \ldots, K\}\,. \tag{1}$$

We model the customer's propensity to move from one state to another using a multinomial logit model:

$$q_{ikk't} = \begin{cases} \dfrac{e^{\mu_{ikk'} + \mathbf{x}^q_{it}\boldsymbol{\delta}_{kk'}}}{1 + \sum\limits_{j=1}^{K-1} e^{\mu_{ikj} + \mathbf{x}^q_{it}\boldsymbol{\delta}_{kj}}} & \text{for } k \in \{1, \ldots, K\}, k' \in \{1, \ldots, K-1\} \\ \dfrac{1}{1 + \sum\limits_{j=1}^{K-1} e^{\mu_{ikj} + \mathbf{x}^q_{it}\boldsymbol{\delta}_{kj}}}, & \text{for } k \in \{1, \ldots, K\}, k' = K\,. \end{cases} \tag{2}$$

The parameter $\mu_{ikk'}$ determines the individual-specific propensity to transition from state k to k', $\mathbf{x}^q_{it}$ is the set of time-varying covariates that might affect the transition process, and the parameter vector $\boldsymbol{\delta}_{kk'}$ captures the effects of those covariates. Let

$$\boldsymbol{\delta} = [\boldsymbol{\delta}'_{11}, \ldots, \boldsymbol{\delta}'_{1(K-1)}, \boldsymbol{\delta}'_{21}, \ldots, \boldsymbol{\delta}'_{2(K-1)}, \ldots, \boldsymbol{\delta}'_{K1}, \ldots, \boldsymbol{\delta}'_{K(K-1)}]'\,.$$

We assume that customers differ in their propensities to transition among the hidden states. These differences in transition probabilities reflect the notion that customers might have different lifetimes or shorter (versus longer) spells of heavy activity. We capture this heterogeneity by allowing the transition parameters to vary across individuals;

$$\mu_{ikk'} = \phi_{kk'} + \eta_{ikk'}\,. \tag{3}$$

where $\phi_{kk'}$ represents the average propensity (across the population) to transit from state k to k', and $\eta_{ikk'}$ represents the individual-level heterogeneity in such behavior.

Let

$$\boldsymbol{\phi} = [\phi_{11}, \ldots, \phi_{1(K-1)}, \phi_{21}, \ldots, \phi_{2(K-1)}, \ldots, \phi_{K1}, \ldots, \phi_{K(K-1)}]',$$
$$\boldsymbol{\eta}_i = [\eta_{i11}, \ldots, \eta_{i1(K-1)}, \eta_{i21}, \ldots, \eta_{i2(K-1)}, \ldots, \eta_{iK1}, \ldots, \eta_{iK(K-1)}]'.$$

Finally, to establish the initial conditions for the relationship states in period 1, we assume that the probability that a customer belongs to state k in period 1 is determined by the vector $\boldsymbol{\pi} = [\pi_1, \pi_2, \ldots, \pi_K]$. We model π_k as

$$\pi_k = \begin{cases} \dfrac{1}{1 + \sum\limits_{k'=2}^{K} e^{\rho_{k'}}}, & k = 1 \\ \dfrac{e^{\rho_k}}{1 + \sum\limits_{k'=2}^{K} e^{\rho_{k'}}}, & k = 2, \ldots, K\,, \end{cases} \tag{4}$$

and let $\boldsymbol{\rho} = \{\rho_k\}_{k=2,\ldots,K}$.

3.2.2 Observed Behaviors

Each period we observe the customer's behavior, which is represented by the random vector $\boldsymbol{Y}_{it} = [\,Y_{it}^o, Y_{it}^c, Y_{it}^u\,]$. Clearly clicking on the content of an email is conditional on opening it; we also assume that the individual must open the email in order to click on an unsubscribe link. We therefore model clicking and unsubscribing behavior conditional on opening. Moreover, conditional on the (latent) state membership, we allow the customer's behavior to be affected by factors exogenous to the customer (e.g., day of the week, firm marketing actions) that might influence behavior without altering the relationship state she occupies at that time. For example, customers might be more prone to open emails on weekdays (versus weekends), or they might be more likely to click on a deal that offers a better discount.

Let $p_{it|k}^o = P(Y_{it}^o = 1 | S_{it} = k)$, the probability that customer i will open an email sent in period t given membership of state k in that period. This is modeled as a function of an underlying propensity to open an email that varies across customers and states (γ_{ik}^o), and customer-level time-varying factors that might affect email opening behavior in a given state ($\mathbf{x}_{it}^o$):

$$p_{it|k}^o = \frac{e^{\gamma_{ik}^o + \mathbf{x}_{it}^o \boldsymbol{\beta}_k^o}}{1 + e^{\gamma_{ik}^o + \mathbf{x}_{it}^o \boldsymbol{\beta}_k^o}}\,. \tag{5}$$

Note that both γ_{ik}^o (capturing the propensity to open) and $\boldsymbol{\beta}_k^o$ (capturing the effect of the covariates) are state-specific. This implies that customers in different states are allowed to have different underlying propensities to open emails and different sensitivity to the time-varying covariates.

Similarly, let $p_{it|k}^c = P(Y_{it}^c = 1|S_{it} = k)$, the probability that customer i will click on at least one link in the email received in period t given membership of state k in that period. Because we model clicking behavior conditional on opening, we have

$$p_{it|k}^c = \begin{cases} \dfrac{e^{\gamma_{ik}^c + \mathbf{x}_{it}^c \boldsymbol{\beta}_k^c}}{1 + e^{\gamma_{ik}^c + \mathbf{x}_{it}^c \boldsymbol{\beta}_k^c}} & \text{if } y_{it}^o = 1 \\ 0 & \text{if } y_{it}^o = 0, \end{cases} \tag{6}$$

where γ_{ik}^c is the propensity to click on at least one deal in the email, and $\boldsymbol{\beta}_k^c$ is the sensitivity to the time-varying covariates ($\mathbf{x}_{it}^c$) that may affect clicking behavior (e.g., characteristics of the deals offered).

We capture the variation in γ_{ik}^o and γ_{ik}^c across individuals and across states in the following manner:

$$\gamma_{ik}^o = \zeta_k^o + \psi_i^o$$
$$\gamma_{ik}^c = \zeta_k^c + \psi_i^c\,,$$

where ζ_k^o and ζ_k^c represent the average propensity (across the population) to open an email and to click on the email content (given opening), respectively, and $\boldsymbol{\psi}_i = \{\psi_i^o, \psi_i^c\}$ represents the individual-level heterogeneity in each of the two behaviors. (Let $\boldsymbol{\zeta} = \{\zeta_k^o, \zeta_k^c\}$.) This splitting into a state-specific component and an individual-specific component ensures identification of the unobserved individual parameters in both the transition probabilities ($\boldsymbol{\eta}_i$) and in the observed behaviors ($\boldsymbol{\psi}_i$) (Ascarza and Hardie 2013).

Finally, let $p_{it|k}^u = P(Y_{it}^u = 1|S_{it} = k)$, the probability that customer i will unsubscribe in period t given membership of state k in that period. This is modeled as a function of a state-specific propensity to unsubscribe (ζ_k^u) and customer-level time-varying factors ($\mathbf{x}_{it}^u$) that might affect the unsubscribing behavior when in a given state.[6] Again, because we model unsubscribing

[6]We do not include unobserved heterogeneity in the propensity to unsubscribe because at most each customer unsubscribes once in the entire observation period, making such a heterogeneity specification empirically unidentified.

conditional on opening, we have

$$p^u_{it|k} = \begin{cases} \dfrac{e^{\zeta^u_k + \mathbf{x}^u_{it}\boldsymbol{\beta}^u_k}}{1 + e^{\zeta^u_k + \mathbf{x}^u_{it}\boldsymbol{\beta}^u_k}} & \text{if } y^o_{it} = 1 \\ 0 & \text{if } y^o_{it} = 0. \end{cases} \tag{7}$$

As with most modeling efforts in contractual settings, we no longer observe customers who unsubscribe.

Note that this specification assumes that the behaviors observed (i.e., opening and clicking) in addition to whether or not the customer has terminated their relationship with the firm are binary in nature. If the data consist of, for example, counts or non-negative continuous quantities, we can easily replace (5) and (6) with state-dependent Poisson or lognormal regressions, respectively. Moreover, our model specification assumes that the three observed behaviors are correlated via the hidden states, but are conditionally independent given state membership. If the researcher were interested in capturing further correlations among the behaviors (e.g., to control for unobserved factors that could shift all behaviors given a relationship state, simultaneously), time-specific random effects could potentially be added to the sub-models of behaviors given state membership.

3.2.3 The Likelihood Function

We now combine the submodels for relationship state evolution and the observed behaviors to derive the likelihood function of the full model. Recall that our model specification assumes that clicking and unsubscribing behavior are conditional on opening. Therefore, conditional on latent state s_{it}, the probability that customer i has observed behavior $\boldsymbol{y}_{it} = [\, y^o_{it}, y^c_{it}, y^u_{it} \,]$ in period t is

$$\begin{aligned} P(\boldsymbol{Y}_{it} = \boldsymbol{y}_{it} | S_{it} = s_{it}) = \mathbb{1}(y^o_{it} = 1) p^o_{it|s_{it}} \Big\{ & \big[\mathbb{1}(y^c_{it} = 1) p^c_{it|s_{it}} + \mathbb{1}(y^c_{it} = 0)(1 - p^c_{it|s_{it}}) \big] \\ & \times \big[\mathbb{1}(y^u_{it} = 1) p^u_{it|s_{it}} + \mathbb{1}(y^u_{it} = 0)(1 - p^u_{it|s_{it}}) \big] \Big\} \\ & + \mathbb{1}(y^o_{it} = 0)(1 - p^o_{it|s_{it}}) \,, \end{aligned} \tag{8}$$

where the indicator function $\mathbb{1}(A)$ equals 1 if A is true and 0 otherwise.

We observe each customer for T_i periods, with the observation period ending either when the customer unsubscribes or at the end of the model calibration period, whichever comes first. The probability of the behavior sequence $\boldsymbol{y}_{i1}, ...\boldsymbol{y}_{iT_i}$ is given by the sum over all possible routes through the (hidden) states the individual could take over time:

$$\mathscr{L}_i(\boldsymbol{\rho}, \boldsymbol{\phi}, \boldsymbol{\delta}, \boldsymbol{\zeta}, \boldsymbol{\beta}, \boldsymbol{\eta}_i\, \boldsymbol{\psi}_i \,|\, \text{data}) = \sum_{s_{i1}=1}^{K} \sum_{s_{i2}=1}^{K} \ldots \sum_{s_{iT_i}=1}^{K} \Bigg\{ P(S_{i1} = s_{i1}) \prod_{t=2}^{T_i} P(S_{it} = s_{it} | S_{i(t-1)} = s_{i(t-1)}) \times \prod_{t=1}^{T_i} P(\boldsymbol{Y}_{it} = \boldsymbol{y}_{it} | S_{it} = s_{it}) \Bigg\}, \tag{9}$$

where $P(S_{i1} = s_{i1})$ is a function of $\boldsymbol{\rho}$, $P(S_{it} = s_{it}|S_{i(t-1)} = s_{i(t-1)})$ is a function of $\boldsymbol{\phi}, \boldsymbol{\delta}$, and $\boldsymbol{\eta_i}$, and $P(\boldsymbol{Y}_{it} = \boldsymbol{y}_{it}|S_{it} = s_{it})$ is a function of $\boldsymbol{\zeta}$, $\boldsymbol{\beta}$ (which contains $\boldsymbol{\beta}^u = \{\boldsymbol{\beta}^u_k\}_{k=1,\ldots,K}$, $\boldsymbol{\beta}^o = \{\boldsymbol{\beta}^o_k\}_{k=1,\ldots,K}$, and $\boldsymbol{\beta}^c = \{\boldsymbol{\beta}^c_k\}_{k=1,\ldots,K}$), and $\boldsymbol{\psi}_i$.

Recalling the notation introduced in (1) and (4), we can write (9) as

$$\mathscr{L}_i(\boldsymbol{\rho}, \boldsymbol{\phi}, \boldsymbol{\delta}, \boldsymbol{\zeta}, \boldsymbol{\beta}, \boldsymbol{\eta}_i\, \boldsymbol{\psi}_i \,|\, \text{data}) = \sum_{s_{i1}=1}^{K} \sum_{s_{i2}=1}^{K} \ldots \sum_{s_{iT_i}=1}^{K} \Bigg\{ \pi_{s_1} \prod_{t=2}^{T_i} q_{is_{i(t-1)}s_{it}t} \prod_{t=1}^{T_i} m_{it|s_{it}} \Bigg\}, \tag{10}$$

where $m_{it|s_{it}} = P(\boldsymbol{Y}_{it} = \boldsymbol{y}_{it}|S_{it} = s_{it})$, as defined in (8).

Following Zucchini and MacDonald (2009), we can rewrite (10) in matrix form as

$$\mathscr{L}_i(\boldsymbol{\rho}, \boldsymbol{\phi}, \boldsymbol{\delta}, \boldsymbol{\zeta}, \boldsymbol{\beta}, \boldsymbol{\eta}_i\, \boldsymbol{\psi}_i \,|\, \text{data}) = \boldsymbol{\pi} \boldsymbol{M_{i1}} \boldsymbol{Q}_{i2} \boldsymbol{M_{i2}} \ldots \boldsymbol{Q}_{iT_i} \boldsymbol{M_{iT_i}} \mathbb{1}'_K\,, \tag{11}$$

where $\boldsymbol{M_{it}} = \text{diag}(m_{it|1}, \ldots m_{it|K})$ and $\mathbb{1}_K$ is a $1 \times K$ vector of ones. We estimate the model parameters using a hierarchical Bayesian framework. (See Web Appendix A for details.)

4 Empirical Application 1: Daily Deals Website

We first apply our model to data from a daily deals website that promotes discounted goods and services to its customers via email. Customers join the service by visiting the company's website and signing up with an email address. At the moment of activation, the customer decides which types of deals she wants to receive (e.g., travel, food) and states her preferred location (e.g., Chicago, New York City). Each day, the company gathers deals from other websites (e.g., Groupon, Gilt, LivingSocial) and sends an email to each subscriber based on the categories she previously selected and her geographical location. The company does not work directly with the retailers and service providers to offer deals in its emails; it simply aggregates deals from different deal websites. As such, the company has limited control and discretion over which deals it decides to offer.

Each email contains up to ten deals and includes pictures of the deals offered as well as other information, such as the identity of the company offering the deal, the discount level, and how long until the deal expires. Once the customer clicks on a deal, she is taken to a different website from which she can purchase that product or service. The company receives a fixed fee every time a customer clicks on a deal. During the period under study, this was the primary source of revenue for the focal firm. Thus, from the perspective of the focal firm, a click on a deal constitutes the financially relevant measure.

4.1 Description of the Data

We obtained data for a set of individuals registered in the New York City area. We focus on the cohort of customers that signed up for the service during February 2012 and track their behavior for 2.5 months. Excluding the days on which the company did not send any emails,[7] this observation window includes 66 periods. (A period is a day on which a customer could have received an email.) The vast majority of the customers in our dataset received an email each period, with the few exceptions (accounting for less than 2% of the customer-periods) occurring when there were no available deals in the selected categories on a particular day. Because the most common method of unsubscribing from the email service is by clicking on the unsubscribe link at the bottom of the email, we only allow a customer to unsubscribe on those days on which an email was received and she opened the email.[8]

Each period the company tracks whether an email was sent to the customer, whether she opened it, which deals she clicked on (if any), and whether she unsubscribed from the service. As more than 90% of the emails were opened on the same day they were sent,[9] we do not model the delay from the date the email was sent to the date it was opened (i.e., if an email was opened, we assume it was opened on the same day in which it was sent). We randomly selected 1,000 individuals who opened at least one email and who did not unsubscribe from the service on the same day they joined (consistent with the firm's definition of a "customer"). Hence, we observe customers for a minimum of 2 and a maximum of 66 periods.

[7]For example, the company did not send emails on Saturdays during our calibration period.

[8]In theory a customer could go to the website and unsubscribe from there. However given that such cases are extremely rare, the company codes such behavior as if the customer unsubscribed from the email sent on that day (or on the day before if no email was sent that day). Once a customer unsubscribes from the service, she stops receiving the emails, and she can no longer access the website to check or buy the available deals; in order to do so, she would need to subscribe again.

[9]This pattern is consistent with that found in previous research (e.g., Bonfrer and Drèze 2009).

In each period, we observe three possible behaviors: opening, clicking, and unsubscribing. In addition, we observe some characteristics of the deals offered, including the size of the discount, where in the email the deal appeared (i.e., its positional order in the email), the source from which the deal could be purchased (e.g., Groupon, LivingSocial), the product category (e.g., food, bar, fitness, travel), and the time left (in days) for the customer to purchase the deal before it expires. Customers tend to select multiple categories; in our sample, 98% of customers chose two or more categories (with an average of 22 and a median of 18). During the period of study, emails included an average of 9.2 deals. This distribution is very concentrated at 10 deals (76.8% of the emails).

4.2 Patterns in the Data

During our observation window, a total of 50,852 emails were sent to the 1,000 randomly selected customers, of which 27.4% were opened, and 5.4% were clicked on at least once. Conditional on being opened, 19.8% of emails were clicked.[10] We observe that customers were heterogeneous in their response to the firm's emails: while some customers almost never opened an email, others opened every email they received. Table 1 shows the summary statistics, across observations and across customers, and Figure 1 shows the histogram of the proportion of emails opened and clicked across customers.

	Mean	Std Dev	Min	Max	N
Across Observations					
% Open	27.4	44.6	--	--	50,852
% Click	5.4	22.7	--	--	50,852
% Click | Open	19.8	39.9	--	--	13,934
Across Customers					
% Open	29.1	28.0	1.5	100.0	1,000
% Click	5.9	8.4	0.0	78.6	1,000
% Click | Open	27.8	30.8	0.0	100.0	1,000

Table 1: Summary statistics for opening and clicking behavior.

In addition to cross-sectional heterogeneity with respect to opening and clicking behavior, customers also vary in their opening and clicking behavior over time. The majority of customers show high levels of activity immediately after joining the service followed by a quick decline in both opening and clicking. These are commonly observed patterns of behavior in noncontractual

[10]These figures are comparable with industry averages for that year (Silverpop 2013).

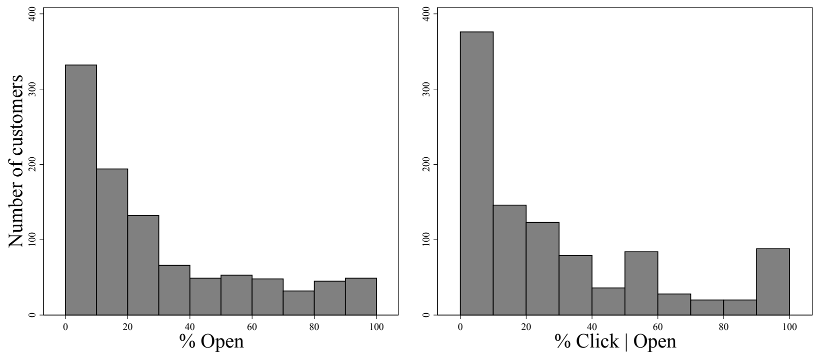

Figure 1: Empirical distributions of % Open and % Click | Open across customers.

settings (e.g., Fader et al. 2010, Schwartz et al. 2014). Figure 2 shows the evolution of the empirical probabilities of opening and clicking from the day the customer joined the firm. Note that while the probability of opening (the leftmost figure) continues declining over time, the probability of clicking given opening (the rightmost figure) stabilizes after approximately 15 periods. The probability of clicking (the middle figure) is simply the product of the other two quantities.

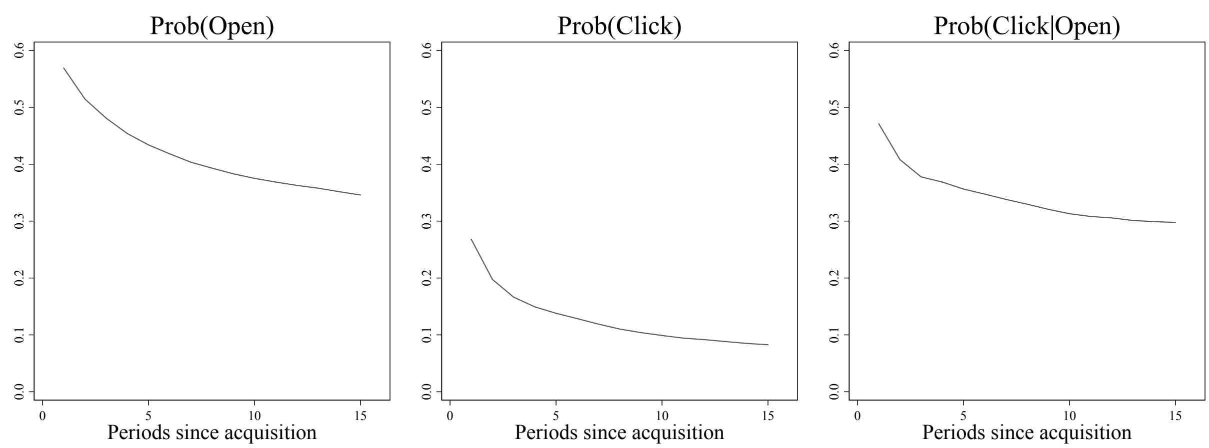

Figure 2: Evolution of the empirical probabilities of opening and clicking.

During our observation window, 8.8% of the customers unsubscribed from the service with the average time to unsubscribing being 22 periods after signing up. The probability of unsubscribing dropped slightly after the first 2–3 periods of joining the service and remained fairly stable thereafter.

In order to explore whether or not unsubscribers differ in their opening and clicking behaviors from those customers who did not unsubscribe, we compute the period-by-period empirical probabilities of opening, clicking, and clicking given opening for each group. So as to ensure that we have enough observations in each group, we focus on those customers who lasted at least 15 periods. If a customer unsubscribed anytime from period 16 to 25 (i.e., within the 10 periods following the 15-period observation window considered above) they are coded as "unsubscribe"; otherwise they are coded as "do not unsubscribe." The period-by-period empirical probabilities for these two groups are plotted in Figure 3.

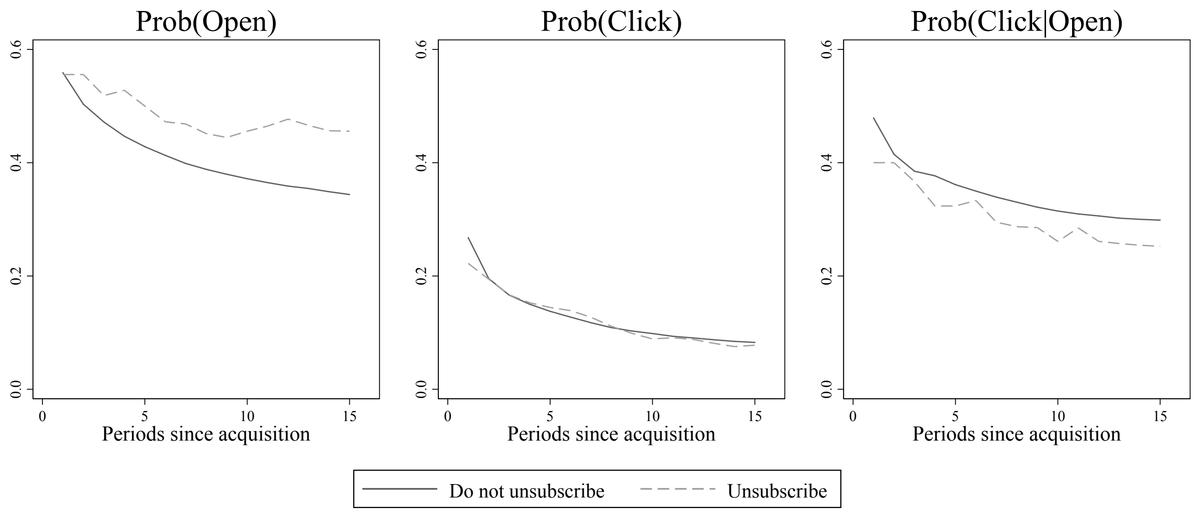

Figure 3: Evolution of the empirical probabilities of opening and clicking, conditional on whether or not the customer unsubscribed in periods 16–25.

Let us make four observations. First, as depicted in the leftmost graph in Figure 3, we see that unsubscribers open more emails than those who stay with the service. This observation is counter to the common finding in contractual settings that churners tend to exhibit lower activity levels prior to churning relative to customers who did not churn (e.g., Ascarza and Hardie 2013, Lemmens and Croux 2006). Second, the opening propensity of customers who did not unsubscribe continues to decline over time. We postulate and later demonstrate that this pattern occurs because the group of customers who did not unsubscribe is composed not only of customers who are satisfied and engaged with the firm but also of a (possibly large) group of customers who have stopped interacting with the firm but have not overtly churned, a pattern likely to occur in hybrid settings. This highlights the need to separate those who did not overtly unsubscribe into latent churners versus active customers. On the other hand, unsubscribers initially show a decline in their propensity to open, but this pattern tapers off and stabilizes soon after, suggesting that unsubscribers keep opening emails until they decide to unsubscribe.

Third, looking at the probability of clicking given opening (the rightmost graph in Figure 3), we see that once the email has been opened, unsubscribers click less than "stayers", suggesting that they do not find the content of the email attractive. Taken together, these two patterns seem to suggest that unsubscribers repeatedly open emails, get disappointed by the content (as reflected by the lack of clicking on emails opened), and eventually unsubscribe. Finally, looking at clicking behavior alone (the middle graph in Figure 3), we see that it is almost impossible to differentiate churners and stayers. Herein lies a word of caution for researchers and analysts working with email marketing: unconditional clicking (or purchasing measures that ignore email opening or website visitation) could be a misleading metric when analyzing churn. It also highlights the need to collect the intermediate measure of opening, and not just clicking behavior.

Thus, the above model-free analyses suggests that treating such hybrid-setting data as coming from a contractual setting, hence ignoring latent churners, inadvertently combines two distinct groups of customers — those who are still active and those who have silently churned — which can result in an erroneous understanding of the firm's customer base and the effects of marketing efforts. Treating such data as purely noncontractual, hence ignoring those who overtly churn, would miss the opportunity for the firm to detect unsubscribing behavior. As we can see, there are meaningful predictors of unsubscribing in our data (e.g., higher opening rates and lower click conditional on opening rates).

Several limitations are associated with the model-free evidence presented in Figure 3. First, we are making inferences about individual-level dynamics from aggregate-level data without controlling for heterogeneity-induced survivor bias. Second, the analysis of dynamics is based on a grouping of customers in terms of their future behavior. Hence, while interesting, the patterns described are only suggestive. Moreover, the model-free analysis does not account for the possibility that the group of stayers could include a group of customers who have silently churned in addition to those customers who are actively engaged with the firm. Our proposed modeling framework overcomes these problems by incorporating dynamics and heterogeneity at the individual level and by measuring the effect of the relevant covariates in each of the processes.

4.3 Model Specification

We now discuss the covariates included in each part of the model. Recall that the main components of the model are the three observed behaviors (open, click, and unsubscribe) and the latent state membership (affecting each of the three behaviors) that is allowed to change over time. While some covariates are only likely to affect customer activity immediately (e.g., customers might be more/less prone to interact with the service during the weekend, or customers are more likely to click on heavily discounted deals), other variables might affect behavior in future periods as well. For example, if a customer is continually exposed to low quality deals, she is not only unlikely to click on that particular deal when she opens the email, but also less likely to open emails in the future. We account for both of these types of effects by incorporating variables that are expected to have a short-term/immediate effect via the state dependent equations (i.e., (5)–(7)) and variables that are likely to have a future impact via the transition equations (i.e., (2)).

4.3.1 State-Dependent Behaviors

Considering the decision of whether or not to open an email, we capture the possible differences in behavior between weekdays and weekends by including a Sunday dummy. We do not include information related to the deals offered in the email as customers cannot see the content of the email before opening it. Therefore, with reference to (5), we have

$$\mathbf{x}_{it}^{o} = [\text{Sunday}]\,.$$

Once a customer has opened an email, the decision to click on any of the deals offered is likely to be affected by the quality of the deals she observes. Rather than estimating the probability of clicking on a particular deal, we estimate the probability of clicking on at least one of the deals contained in the email[11] and incorporate in the clicking model a measure of the overall quality of the deals contained in the email.

We assume that each deal provides the user with the utility $\mathbf{z}_{ijt}\boldsymbol{\lambda}_k^c + \epsilon_{ijt}$, where $\mathbf{z}_{ijt}$ includes the characteristics of deal j offered to customer i at time t, $\boldsymbol{\lambda}_k^c$ captures customer preferences over those characteristics, given that the customer belongs to state k, and ϵ_{ijt} is a zero mean stochastic shock.

[11] While a customer can click on multiple deals in any given email, such behavior is rare. Of the emails clicked, 88% had one deal clicked, 9.51% two deals clicked, and 2.4% three or more deals clicked. We estimated a model in which we allow the customer to click on more than one deal (using a shifted-Poisson distribution to characterize the number of deals clicked). The fit of this model specification was inferior to that of our proposed model specification.

Note that we allow customers in different states to derive different value from different aspects of the deals. For example, customers at risk of overtly churning may pay more attention to the degree of discount offered in the deal. Our measure of the overall quality of deals is the expected utility across the N_{it} deals contained in the email received by customer i in period t, computed as $\sum_{j=1}^{N_{it}} \mathbf{z}_{ijt}\boldsymbol{\lambda}_k^c$.

For each deal we observe the following six attributes: *discount* applied by the vendor (coded as the proportion of the price that has been discounted), the number of days until the deal expires (denoted as *time left*), whether the deal includes a *food*- or *fitness*-related product/service (coded as dummy variables), the log of the positional *order* in which the deal was shown to the customer (where 1 corresponds to the deal shown at the top of the email, 2 to the second deal, and so on), and the vendor from which the deal was collected (e.g., LivingSocial, Groupon). To keep the number of parameters manageable, we represent the 54 different vendors in our data via a continuous measure, *source*, that captures the prevalence of a particular vendor (i.e., how often the vendor appears in the deals' website emails). We operationalize source as the logarithm of the number of times each vendor was offered by the focal company, standardized across vendors. Table 2 reports the summary statistics for the deal characteristics variables.

	Mean	Std Dev	Min	Max	N
Across Deals					
Discount	0.58	0.14	0.00	1.00	3,684
Time left for purchase (in days)	4.07	3.50	0.00	64.38	3,684
Food dummy	0.20	0.40	--	--	3,684
Fitness dummy	0.06	0.24	--	--	3,684
Source	1.07	0.63	−2.68	1.57	3,684
Across Observations					
Order (within email)	5.11	2.83	1.00	10.00	419,214

Table 2: Summary of deals' characteristics included in the model.

In addition to considering the expected utility of the deals offered in each email, we control for two other factors that could influence the state-specific probability of clicking: the day of the week on which the email was sent (using a Sunday dummy as above) and the size of the email (using the log of the number of deals in the email). We therefore rewrite (6) for this empirical application as:

$$p^c_{it|k} = \begin{cases} \dfrac{e^{\gamma^c_{ik} + \mathbf{x}^c_{it}\boldsymbol{\beta}^c_k + \sum\limits_{j=1}^{N_{it}} \mathbf{z}_{ijt}\boldsymbol{\lambda}^c_k}}{1 + e^{\gamma^c_{ik} + \mathbf{x}^c_{it}\boldsymbol{\beta}^c_k + \sum\limits_{j=1}^{N_{it}} \mathbf{z}_{ijt}\boldsymbol{\lambda}^c_k}} & \text{if } y^o_{it} = 1 \\ 0 & \text{if } y^o_{it} = 0, \end{cases} \tag{12}$$

where

$$\begin{aligned} \mathbf{x}^c_{it} &= [\text{Sunday}, \log(\#\text{deals})], \text{ and} \\ \mathbf{z}_{itj} &= [\text{Discount}, \text{Time left}, \text{Food}, \text{Fitness}, \text{Source}, \text{Order}]. \end{aligned}$$

Finally, regarding unsubscribing behavior, we would expect it to be affected by the overall nature of the deals shown in the email. We capture this by including the average characteristics of all the deals contained in the email received that period. We also allow the probability of unsubscribing on a particular day to be affected by whether or not the deal was offered on Sunday and by the log of the number of deals offered in the email received that day. Therefore, with reference to (7), we have

$$\mathbf{x}^u_{it} = [\text{Avg. Discount}, \text{Avg. Time left}, \%\text{Food}, \%\text{Fitness}, \text{Avg. Source}, \text{Sunday}, \log(\#\text{deals})].$$

4.3.2 State transitions

As previously discussed, while some variables (e.g., day of the week) are likely to shift behavior only in the short-run, other variables might also affect future actions. For example, a customer who has repeatedly been exposed to low quality deals may be more likely to transition to a latent state associated with a higher probability of overtly churning, or she might be less prone to open emails in the future, silently leaving the firm. While we could incorporate the effect of the average characteristics of all attributes in the deals, as we did in the unsubscribing submodel, such a specification would imply $K(K-1)$ parameters per deal attribute variable, with K being the number of states. We therefore opt for a more parsimonious measure of overall deal quality.

We model such an effect by incorporating a "stock" variable, which captures the effect of the overall quality of the deals the customer has been exposed to, in the state transition equation. We create this variable in the following manner. First, we create a measure of deal "popularity" by computing the proportion of customers (from outside our sample) exposed to each deal who clicked on the deal (given opening).[12] For each email a customer receives, we compute its overall quality (Quality_{it}) by taking the average of the "popularity" of all the deals contained in the email.[13] Finally, we create the quality stock variable that captures the effect of the quality of the current and previous emails, by using an exponential smoothing function, similar to the Adstock variable commonly used to measure advertising carryover (e.g., Danaher, et al. 2008),

$$\text{QualStock}_{it} = \begin{cases} (1-\varphi)\text{Quality}_{it} + \varphi\text{QualStock}_{i(t-1)} & \text{if } y^o_{i(t-1)} = 1 \\ \varphi\text{QualStock}_{i(t-1)} & \text{if } y^o_{i(t-1)} = 0 \end{cases} \tag{13}$$

where Quality_{it} is the average quality (as measured by popularity) of the deals shown in the email sent to customer i in period t, and φ measures the level of "memory."[14] Because the content of an email in period t affects the transition among states between periods t and $t+1$, we include the one-period lag of this stock variable ($\text{QualStock}_{i(t-1)}$) in the transition equations for period t.

In addition to the quality stock variable, we control for the rare occasions in which a customer did not receive an email on a particular day, by incorporating a "Number of periods since last email" variable in the transition equations. Therefore, with reference to (2), we have

$$\mathbf{x}^q_{it} = [\text{Lag(QualStock)}, \text{Number of periods since last email}].$$

4.4 Selecting the Number of States and Model Fit

We split the 2.5 months of data into a calibration period (from February 1 to March 15) and a validation period (from March 16 to April 15). We estimate the model varying the number of states from one to four, and compute the log predictive density, the Watanabe-Akaike Information Criterion (WAIC), and the mean square error (MSE) and mean absolute percentage error (MAPE)

[12] We only consider opened emails as it is impossible for the customer to observe the quality of the offered deals without opening the email that contains the deals.

[13] We standardize the quality variable such that $\text{Quality}_{it} = 0$ corresponds to an email of average quality.

[14] We estimate φ using a grid search, with points 0.2, 0.4, 0.6 and 0.8. The results presented in Section 4.5 are associated with $\varphi = 0.2$, which provide a better fit to the data. We initialize the stock variable at zero, implying that $\text{QualStock}_{it} = 0$ if the customer has not opened any email unto and including period t. In other words, we initialize the quality stock variable at the level of an average email.

between the observed and predicted opening, clicking and unsubscribing behaviors.[15,16] With reference to Table 3, the model with the best log predictive density, WAIC and MAPE is the model with three hidden states, whereas the model with four states provides slightly lower MSE. Of these two, we chose the more parsimonious specification with three states. In addition to the metrics shown in Table 3, we also compute alternative WAIC and DIC (as suggested in Gelman et al. 2013), all favoring the model with three states. (See Web Appendix B1 for the results of the two-state and four-state specification and a discussion of the differences and similarities.)

While we did not use the holdout sample for model selection purposes, we report the (out-of-sample) predictive performance of each specification. Specifically, we report the MSE, which aggregates prediction error across all three behaviors, and the area under the curve (AUC), which measures the accuracy of identifying customers who are more likely to overtly churn. In terms of out-of-sample prediction, the model with four states has the lowest MSE while the three-state model has the best AUC.

	In-sample				Out-of-sample	
# States	Log pred. density	WAIC	MSE	MAPE	MSE	AUC
1	−12,766	25,881	0.188	37.4	0.182	0.575
2	−12,652	25,731	0.165	32.4	0.130	0.637
3	−12,537	25,665	0.161	26.9	0.123	0.720
4	−12,611	25,804	0.160	29.5	0.114	0.688

Table 3: Model fit measures as a function of the number of states.

4.5 Results

We start by summarizing the state-specific behavior for each latent state as well as the transitions among the latent states, with the goal of characterizing the latent states. This is followed by a discussion of the impact of the covariates on customer behavior in each of the states. We then use the model parameters to provide insights that would help the firm better manage its customer base and increase engagement (and hence revenue) among its customers. Finally, we show how accurately

[15] Log predictive density (Gelman et al. 2013, p. 167) is computed as the posterior mean of the log-likelihood function evaluated in each draw of the Gibbs sampler.

[16] For each individual in each period for each of the three behaviors, squared error is computed as the square of the difference between the probability of the behavior and the actual behavior (0/1). The mean is computed across individuals, periods, and the three behaviors. APE is computed at the period level and averaged across periods to give us MAPE. Specifically, we compute the absolute percentage error in predicting the total number of opens, clicks, and unsubscribers on a given day and then average across the three behaviors.

the model forecasts behavior outside the calibration sample and compare its performance to that of several benchmark models.

4.5.1 Characterizing the Latent States

Table 4 presents the posterior means and 95% central posterior intervals (CPIs) of the probabilities of opening, clicking, and unsubscribing, for each of the hidden states. For ease of interpretation, we report probabilities instead of the underlying model parameters. These probabilities were computed for a typical day, defined as a weekday in which an email with "average" deals was sent.

		Posterior mean	95% CPI	
Prob(Open)	State 1	0.561	0.526	0.601
	State 2	0.520	0.478	0.554
	State 3	0.053	0.041	0.066
Prob(Click)	State 1	0.051	0.023	0.087
	State 2	0.234	0.159	0.321
	State 3	0.011	0.004	0.024
Prob(Unsubscribe)	State 1	0.005	0.003	0.008
	State 2	0.001	0.001	0.002
	State 3	0.001	0.000	0.001

Table 4: Posterior means and 95% CPIs of the state-dependent probabilities of opening, clicking, and unsubscribing, for an email sent on a weekday with "average" deal characteristics.

Customers in state 2 appear to be the most active customers. Upon receiving an email, the mean probability that a customer in state 2 opens it is 0.520, and the probability of clicking on at least one deal is 0.234. These propensities are much higher than the simple averages across the entire customer base reported in Table 1. These are the most profitable customers for the company as they have the highest clicking rates. At the other extreme, customers in state 3 rarely open an email (Prob(Open) = 0.053) and rarely unsubscribe from the service (Prob(Unsubscribe) = 0.001). These customers seem to have the minimal level of interaction with the firm. Customers in state 1 exhibit interesting behavior, as evidenced by their high propensity of opening emails (Prob(Open) = 0.561) but a low propensity to click on the same emails. The propensity of customers in state 1 to unsubscribe from the service is five times that of customers in states 2 or 3. Note that the 95% CPIs for the state 1 unsubscribe probabilities do not overlap with those of the other states, implying that customers in state 1 have a significantly higher propensity to unsubscribe than customers in either of the other two states.

Dynamics in customer behavior are captured by allowing customers to transition among the latent states. We compute the posterior distribution of the elements of the transition matrix for each individual (i.e., $\boldsymbol{Q}_{it}$). (For ease of interpretation, we compute $\boldsymbol{Q}_{it}$ for a typical day in which the Lag(QualStock) variable corresponds to the average in the data and the number of periods since last email is set to 1.) We then compute the posterior means of these quantities for each individual and report the average and the 95% heterogeneity interval across all individuals.[17] These quantities are reported in Table 5 and should be interpreted in the following manner. With respect to the middle two rows, an "average" individual in state 2 (the high activity state) has a 0.728 probability of remaining in that state in the next period, a 0.099 probability of switching to state 1, and a 0.173 probability of switching to state 3. Individuals vary substantially in their propensity to transition between states, as evidenced by the 95% heterogeneity intervals.

	To state		
From state	1	2	3
1	0.503	0.185	0.312
	[0.049 , 0.994]	[0.000 , 0.576]	[0.006 , 0.525]
2	0.099	0.728	0.173
	[0.024,0.168]	[0.544 , 0.943]	[0.031 , 0.301]
3	0.040	0.006	0.954
	[0.001,0.122]	[0.001 , 0.014]	[0.865 , 0.998]

Table 5: Average and 95% heterogeneity interval (reported in square brackets) of the individual posterior means of the state transition probabilities.

We observe that state 3 is very "sticky." That is, once a customer enters this low activity state, it is very unlikely that she will move to a different state. Also note that there is very little heterogeneity on that dimension, with 95% of the sample having a probability of staying in state 3 between 0.865 and 0.998. State 1, on the other hand, is less sticky with only a 50.3% chance of staying in the same state in the next period; however this probability is very heterogeneous across customers.

[17] Note that the 95% heterogeneity interval captures heterogeneity of the transition probabilities across individuals rather than the parameters' precision.

Finally, Table 6 shows the posterior probabilities for the initial conditions for the relationship states in period 1. Not surprisingly, and given that we analyzed customer behavior since the moment they joined the company, the large majority of customers started their relationship with the company in the state with the highest levels of opening and clicking activity (state 2).

	Posterior mean	95% CPI	
State 1	0.337	0.217	0.477
State 2	0.657	0.509	0.781
State 3	0.006	0.001	0.018

Table 6: Posterior means and 95% CPIs of the initial probabilities of belonging to each of the states in period 1.

Combining the insights from the results presented in Tables 4 – 6, we can now label the states. State 2 is characterized by a high propensity to both open and click. Accordingly, we name this the "engaged" state. State 3 is characterized by a very sticky, almost absorbing, state with low propensities to both open and unsubscribe. Thus, this state closely resembles the "dead" state of traditional noncontractual-setting models such as the Pareto/NBD, BG/NBD, and BG/BB. We label this the "silently gone" state. Note that in contrast to these models, we have not forced an absorbing dead state on the data but have rather recovered it from the data. Finally, state 1 is characterized by a high propensity to open an email but the lowest propensity to click given that an email was opened, and interestingly, the highest propensity to unsubscribe. The pattern of behavior of these customers is similar to that reported by the unsubscribers group in Figure 3. Thus, customers in this state are not disengaged with the firm — after all, they are opening emails — but they do not seem to find what they are looking for, which puts them at a higher risk of unsubscribing. We therefore name this the "at risk" state.

4.5.2 Covariate Effects

We now discuss how changes in the firm's product offering as well as external factors affect customer behavior, both in the short term (affecting the state-dependent behaviors) and in the longer term (affecting the transition probabilities).

Covariates effects on the state-dependent behaviors

Table 7 reports the posterior means and 95% CPIs for the effects of the covariates on each behavior. Recall from Section 3 that we allow the covariate effects to be state specific. Hence, the effects of the same covariate might vary across columns, depending on which state the customer occupies. Regarding opening behavior, "at risk" customers are more likely to open emails on Sundays than on weekdays, whereas the "engaged" customers are more likely to open emails on weekdays than on Sundays.

Regarding clicking behavior, we also find that customers in different states exhibit different responses to the covariates. Customers in the "engaged" state are sensitive to the characteristics of the featured deals. Interestingly, preferences for some deal characteristics vary across states. For example, both "at risk" and "engaged" customers click significantly more on deals with greater discounts; the magnitude of such effect is much larger for customers in the "engaged" state than for those in the "at risk" state (0.12 versus 0.29). Customers in the "engaged" state are also more likely to click as the number of deals in an email increases, whereas customers in the other two states are not significantly affected by the number of deals included in the email. These differences in sensitivity to deal characteristics are important for managers interested in tailoring the content of an email to specific segments in order to increase customer profitability. We explore different options as to how to alter the content of the email and its consequences in Section 4.5.4.

Finally, "engaged" customers are less likely to unsubscribe from emails that include deals from popular sources, whereas the other two states do not show significant differences in unsubscribing behavior.

Covariate effects on the state transition probabilities

We now describe the effect of past email content, as measured by the stock variable, on the (latent) state transition probabilities. For ease of exposition, we present the magnitude of these effects by reporting the average (across individuals) of the posterior means of each individual's transition probabilities for different levels of the quality stock variable.[18] The "default" case corresponds to the transition matrix shown in Table 5 in which we use the average value for the quality stock variable. We operationalize high ("better deals") and low ("worse deals") levels of Lag(QualStock)

[18]The parameters associated with (2) are reported in Table B1, Web Appendix B.

	State		
	"At Risk"	"Engaged"	"Silently gone"
Effect on Opening			
Sunday	**0.70**	**−0.56**	**0.35**
	[0.16 , 1.25]	[−0.88 , −0.20]	[0.05 , 0.62]
Effect on Clicking, given Opening			
Sunday	0.43	**−0.45**	0.02
	[−0.07 , 0.91]	[−0.82 , −0.07]	[−0.63 , 0.69]
log(#deals)	0.17	**0.63**	0.32
	[−0.29 , 0.69]	[0.21 , 1.06]	[−0.48 , 1.22]
Discount	**0.12**	**0.29**	**−0.13**
	[0.00 , 0.32]	[0.10 , 0.55]	[−0.33 , 0.00]
Time left	**−0.01**	**0.01**	**−0.02**
	[−0.03 , 0.00]	[0.00 , 0.02]	[−0.06 , 0.00]
Food	**−0.12**	**−0.12**	**−0.10**
	[−0.27 , 0.00]	[−0.21 , −0.04]	[−0.24 , 0.00]
Fitness	0.02	**0.26**	−0.05
	[−0.06 , 0.13]	[0.01 , 0.44]	[−0.15 , 0.02]
Source	−0.03	0.03	0.03
	[−0.11 , 0.02]	[−0.03 , 0.10]	[−0.03 , 0.14]
Order	−0.02	−0.01	−0.03
	[−0.07 , 0.02]	[−0.07 , 0.03]	[−0.12 , 0.02]
Effect on Unsubscribing, given Opening			
Avg. Discount	0.63	−0.41	0.72
	[−0.427 , 1.799]	[−1.64 , 0.69]	[−0.53 , 2.00]
Avg. Time left	−0.09	−0.03	0.02
	[−0.30 , 0.06]	[−0.19 , 0.09]	[−0.14 , 0.15]
%Food	−0.24	0.91	0.92
	[−1.27 , 0.71]	[−0.20 , 2.01]	[−0.31 , 1.94]
%Fitness	−0.17	0.21	−0.88
	[−1.59 , 1.24]	[−1.07 , 1.51]	[−2.37 , 0.58]
Avg. Source	0.82	**−1.15**	−0.26
	[−0.04 , 1.75]	[−2.03 , −0.30]	[−1.53 , 1.04]
Sunday	**−1.48**	0.66	**−1.51**
	[−2.54 , −0.45]	[−0.35 , 1.60]	[−2.65 , −0.46]
log(#deals)	0.44	−0.71	−0.13
	[−0.21 , 1.09]	[−1.44 , 0.14]	[−0.95 , 0.78]

Table 7: Posterior means of the effect of the covariates on the state-dependent probabilities. Numbers in bold are associated with 95% CPIs (in brackets) not including 0.

as +/− one standard deviation around the average in the data. The transition probabilities for each of the three scenarios are reported in Table 8.

	Default			Better deals			Worse deals		
	1	2	3	1	2	3	1	2	3
1 (“at risk”)	0.503	0.185	0.312	0.471	0.226	0.303	0.535	0.148	0.316
2 (“engaged”)	0.099	0.728	0.173	0.106	0.742	0.152	0.092	0.712	0.195
3 (“silently gone”)	0.040	0.006	0.954	0.030	0.007	0.963	0.054	0.005	0.942

Table 8: Average of the individual posterior means of the transition probabilities for different levels of prior deal quality.

First, consider the average impact of the quality of past emails for customers belonging to the “at risk” state (first row of Table 8). Exposure to better deals makes “at risk” customers more likely to transition to the “engaged” state. (The average transition probability increasing from 0.185 to 0.226 is statistically significant at the 5% level.) This shows that, even though we did not find an immediate impact of the deal characteristic on unsubscribing behavior (no significant effects on the first column of Table 7), sending better deals to “at risk” customers can prevent unsubscribing (in future occasions) as better deals can move customers from the “at risk” state to the “engaged” state, where the probability of unsubscribing is notably lower. Regarding the already “engaged” customers (second row of the transition matrix), these effects are less pronounced. Nevertheless, we find that being exposed to better deals reduces the average probability of moving from “engaged” to “silently gone” (with the transition probability decreasing from 0.173 to 0.152). Finally, for the customers that already are in the “silently gone” state, we find a very weak, almost negligible, impact of offering better deals.

The effect of the deal quality stock variable on the state transition probabilities provides further support for our characterization of the latent states in three primary ways. First, the quality of emails influences a customer’s propensity to move from the “at risk” to the “engaged” state. This finding corroborates our first assertion that the “at risk” state does not represent customers who are just not interested in purchasing (i.e., browsers who would never click) but rather captures those who were not able to find what they wanted. Second, customers are less likely to transition from the “engaged” to the “silently gone” state when they receive emails with better deals, providing further evidence that the “engaged” and “silently gone” states are, indeed, states, not fixed underlying customer traits. Finally, customers in the “silently gone” state are not affected by changes in the content of the emails. This effect is not surprising as this state consists of customers who almost never open their emails.

4.5.3 Relationship State Recovery

We can also use the parameter estimates to compute the probability of each customer belonging to each state at any time period. We use the smoothing approach (Zucchini and McDonald 2009) to calculate the probability that customer i is in state k in period t. The smoothing probability is given by

$$P(S_{it} = k \,|\, \boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{iT_i}) = \boldsymbol{\pi} \boldsymbol{M}_{i1} \boldsymbol{Q}_{i2} \boldsymbol{M}_{i2} \cdots \boldsymbol{q}_{it}^{\cdot k} m_{it|k} \boldsymbol{q}_{i(t+1)}^{k \cdot} \cdots \boldsymbol{Q}_{iT_i} \boldsymbol{M}_{iT_i} \mathbb{1}_K' / \mathscr{L}_i \,, \tag{14}$$

where $\boldsymbol{q}_{it}^{\cdot k}$ is the kth column of the transition matrix $\boldsymbol{Q}_{it}$, $m_{it|k}$ is the kth diagonal element of $\boldsymbol{M}_{it}$, $\boldsymbol{q}_{it}^{k\cdot}$ is the kth row of $\boldsymbol{Q}_{it}$, and $\mathscr{L}_i$ is the likelihood of the observed sequence of customer i's behavior, as given in (11). Averaging these probabilities across customers gives us an estimate of the proportion of the customer base in each state at any time period. We plot in Figure 4 the evolution of these proportions over time since joining the service. Figure 4 also shows the proportion of customers who have overtly churned up to any given period (represented by the top bar).

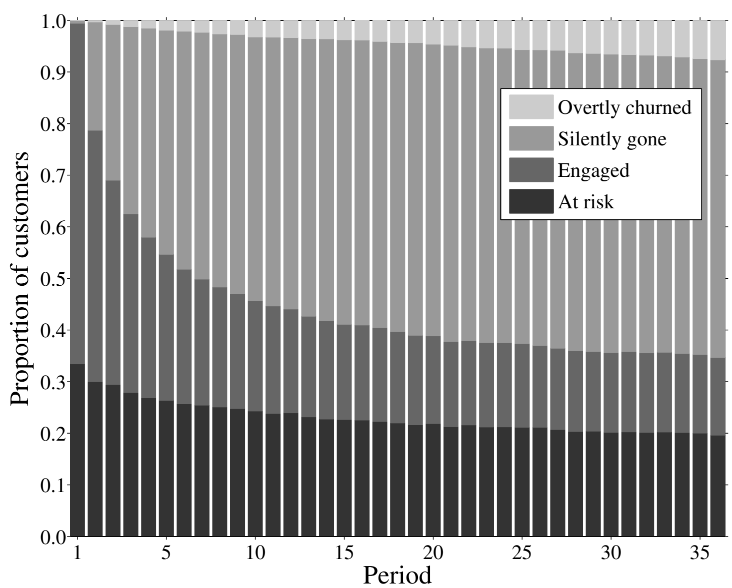

Figure 4: The evolution of the proportion of customers in each state as a function of time since joining the service.

Understanding the evolution of customers' state membership could help the manager better understand her customer base. For example, we find that the number of customers in the "engaged" state decreases during the first few periods, then stabilizes to less than 20% of the customer base. Similarly, while the share of "at risk" customers also decreases, it does so at a much slower rate relative to the "engaged" customers. It is interesting to note that even though on average a customer has only a 50% chance of remaining in the "at risk" state (recall the average transition probabilities from Table 5), there is a high degree of heterogeneity in that probability, with a stable group of customers who remain "at risk" for a long period of time. This pattern suggests that many customers might be at risk for quite a few periods before unsubscribing, which may offer the firm a window of opportunity to react.

The remaining groups capture the proportion of customers churning — either overtly or silently — from the service. The top bar captures the proportion of customers who have already unsubscribed (i.e., 'overtly churned') from the service by each period. Consistent with the nature of contractual relationships (i.e., once a customer unsubscribes, she is no longer a customer), and given that we analyze a cohort of customers, the size of this group is increasing overtime. Unlike the overt churners, "silently gone" customers (second-to-top bar in Figure 4) are part of our data (as they keep receiving communications from the company) and are captured by one of the latent states. Consistent with the nearly absorbing nature of that latent state, the size of the "silently gone" group also increases in size overtime, even more rapidly than the overt churners. It is apparent from the figure that customers of this service are more likely to leave silently than overtly.

Figure 4 provides an illustration of how the model captures latent attrition while allowing the "still alive" customers to belong to different groups that evolve over time. Because the "silently gone" population is increasing over time, modeling our setting as a contractual setting — thus grouping together all customers who did not overtly churn — will significantly overstate the size of the truly active customer population. Furthermore, these insights can help marketers avoid keeping customers who are long gone in their communications database. This is especially relevant in email marketing settings because firms are increasingly interested in removing "dead" customers from their mailing lists so as to maximize the overall deliverability of their email communications.[19] Most importantly, combining these insights with those obtained in Section 4.5.2, our model offers clear guidance as to which marketing actions should be undertaken to increase activity among existing customers both in the short-run (through increasing clicking behavior immediately), as well as in future periods (through changing the mix of customers via the latent states). In the next section we explore the options that the focal firm could implement (given their business model) to better manage its customer base and increase profitability.

[19] Amongst other things, mailbox providers use recipients' engagement with emails (email opening and clicking) from each sender to determine whether to re-route emails from the sender to the recipients' junk folder.

4.5.4 Actively Managing the Customer Base

Putting all the pieces together, our model identifies three latent states of behavior, which we characterize as "at risk," "engaged," and "silently gone." Customers in the "at risk" state open emails frequently but rarely click on any deals. Their probability of clicking on a deal is higher when deals offer greater discounts, yet their overall probability of clicking is very low. Customers in the "engaged" state are active email openers who also have high propensity to click on deals. They are more likely to click on deals that offer high discounts, that come from popular sources, or that have more days before expiring. They tend to prefer fitness-related deals and click less when the deal is food-related. In addition to having different clicking patterns, "engaged" customers also differ from "at risk" customers in their significantly lower propensity to unsubscribe (Table 4). Interestingly, the quality of deals that customers are exposed to has an effect above and beyond the immediate actions of opening, clicking, and unsubscribing. We find that by sending emails with better deals, "at risk" customers have a higher chance of becoming "engaged," which determines their behavior in future periods. Finally, those in the "silently gone" state do not exhibit any sign of activity, and are very unlikely to change, as reflected by their negligible reaction to the email content and their nearly zero probability of transitioning to a different state.

A natural question to ask is whether the firm can use these insights to increase the activity and profitability of the existing customer base. To investigate this question, we conduct a set of counterfactual analyses, simulating customer behavior while varying the nature of the deals offered by the company to its customers. Recall that the focal company does not work directly with the vendors offering these deals but collects deals from other websites. Accordingly, the focal firm cannot alter the characteristics of the deals, but it can choose *which* deals to offer. Furthermore, given the large supply of deals in our context, the company can easily improve the deal offering by better matching deals to customers with the objective of increasing clicking behavior and reducing customers' probabilities of churning, either silently or overtly. Therefore, in our counterfactual analyses we measure the impact of improving the deal offerings on opening, clicking, and unsubscribing behavior, as well as on reducing latent attrition by preventing customers from transitioning to the "silently gone" state.

We select customers who had not unsubscribed by the end of the calibration period and predict their behavior under three different scenarios: (i) *Status quo*, which assumes the company provides the same deals as we observe in the data, (ii) *Better deals*, which assumes that the company selects the most popular deals and sends them to every subscriber, and (iii) *Personalized emails*, which assumes the company personalizes the emails based on the (hidden) state the customer belongs to at the end of the calibration period. We determine the hidden state to which each customer belongs by first calculating her state membership probabilities using (14) and then assigning her to a state if the probability of belonging to that state is greater than 0.8.[20,21]

The way we operationalize the second and third scenarios is by selecting deals from the pool of deals we observe in the data. (This mimics the way the company operates by selecting deals from those available in the market.) More specifically, the deals for the second scenario are drawn from the pool of deals with the highest popularity, as described in Section 4.3. For scenario three, we rely on the parameter estimates from Table 7 and create three pools of deals, one for each hidden state. The "at risk" pool contains deals that offer a high discount, do not offer food-related products, and have lower number of days to be purchased. The "engaged" pool also contains high discount deals, but the deals are more likely to be for fitness services, and have more days left for the deal to be purchased. Moreover, because "engaged" customers prefer more (rather than fewer) deals, engaged customers in scenario three are offered 15 deals instead of 10. Finally, the "silently gone" pool contains shorter, lower discount deals that are not food related.

Figure 5 depicts the impact of each of the scenarios on customers in each of the latent states in terms of their observed behaviors (opening, clicking and unsubscribing) over 12 periods (i.e., two weeks excluding Saturdays). The left-most column shows the effect of the company actions on customers who are in the "at risk" state at the end of the calibration period, the middle column on those in the "engaged" state, and right-most column on those in the "silently gone" state. The y-axes represent the difference in the probability of the behavior (open, click, and unsubscribe in rows 1, 2, and 3, respectively) between the default (scenario one) and each of the other scenarios in which the company is improving the quality of its offerings. (0 implies no change.)

Let us start by analyzing clicking (second row), as this behavior is directly linked to the company's bottom line. Not surprisingly, the company can increase clicks, hence profits, among its current customers by choosing the deals it will offer in a more selective manner. By simply offering better deals (scenario 2), click rates increase by one percentage point among "at risk" customers

[20]Our use of such a high cutoff means we are selecting customers for whom we are fairly certain as to which state she belongs to. If a customer has no posterior probability greater than 0.8, we discard her from this analysis.

[21]For a customer who is not part of the calibration sample, state membership probabilities can be calculated by first computing her posterior distribution (given her observations and the population-level parameters from the calibration sample) and then using the resulting individual-level parameter values to evaluate (14).

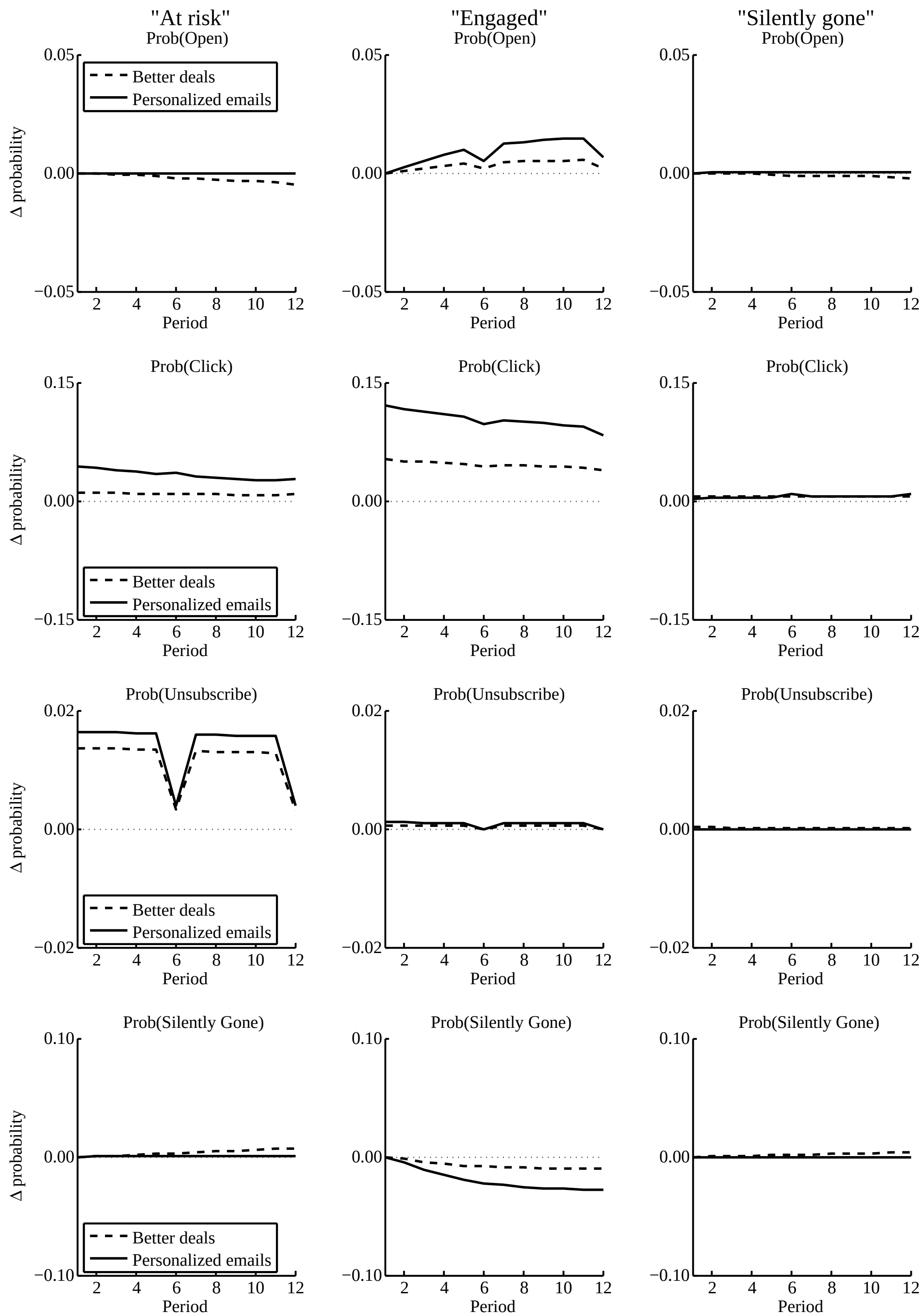

Figure 5: Differences in opening, clicking, unsubscribing and silent churn probabilities (relative to status quo) as the company alters the quality of the emails.

and by five percentage points among "engaged" customers. The increase in click rates is significantly larger when the company leverages the insights from the model regarding the customers' state membership, and tailors the deal offered based on the hidden states customers belong to (scenario 3). In this case, the increase in click rates are 5 and 10 percentage points for customers in the "at risk" and "engaged" states, respectively (left-most figures of the second row of Figure 5).

The effects on opening behavior are less obvious because the impact of the policies on opening behavior takes time to pick up. This is because improving the quality of deals cannot impact opening directly (as customers need to open the email first to see the quality of the deals offered) but affects future behavior by altering the hidden states customer belong to. We find that improving the quality of the deals offered increases opening rates among "engaged" customers over time. This pattern is driven by the impact of the policies on the probability of moving to the "silently gone" state (last row of Figure 5). Indeed we find that improving the quality of the deals reduces the probabilty of "engaged" customers of silently leaving the firm.

Finally, consistent with our previous results and with our characterization of state 3 (the "silently gone" state), we find that improving the product offering has little to no effect on customers who are "silently gone." These customers rarely interact with the service, meaning they would not be aware of the changes in quality of the deals offered. The firm would need to find other means of "saving" these customers, to the extent that doing so is at all possible.

4.6 Assessing Predictive Performance Relative to Benchmark Models

We now compare the out-of-sample predictions (over the one-month validation period) generated by our model with those generated using both simpler versions of our model specification and more traditional approaches that model behavior as a function of past behavior and/or lagged covariates. In selecting our benchmark models, we choose models commonly used to predict (overt) churn in contractual settings while controlling for the effect of past levels of activity.[22] Accordingly, our benchmark models include models that remove the latent dynamics offered by the HMM and replace them with observed dynamics using lagged covariates and/or the recency and frequency measures. Such models are constructed by estimating a single-state HMM and augmenting the vectors of covariates $\mathbf{x}_{it}^o$, $\mathbf{x}_{it}^c$, and $\mathbf{x}_{it}^u$ in (5)–(7). (Further details of the model specifications and the

[22] Comparing our models to those commonly used to predict churn in contractual settings gives us a sense of the (potential) loss in predictive ability due to the failure to separate overt churners from latent churners. Note that we do not compare the predictive ability of our model to that of the standard models for noncontractual settings, as such models are not able to predict unsubscribing behavior. Furthermore, such models typically model one behavior at a time as opposed to multiple behaviors (e.g., opening, clicking, and unsubscribing).

full set of results are provided in Web Appendix B3.) Additionally, we estimate a static benchmark version of our HMM in which transitions between states are not allowed.

We consider four benchmark models:

- *Lagged covariates*, in which the opening, clicking, and unsubscribing behaviors are modeled as a function of $\mathbf{x}_{it}^u$, $\mathbf{x}_{it}^o$, and $\mathbf{x}_{it}^c$, augmented by $\mathbf{x}_{it}^q$, as well as the lagged values of these covariates.

- *RF (no covariates)*, in which the opening, clicking, and unsubscribing behaviors are modeled as a function of four recency-frequency variables: rec_{it}^o, freq_{it}^o, rec_{it}^c, and freq_{it}^c.[23] We specify the recency and frequency variables in the following manner. freq_{it}^o is computed as the proportion of emails customer i has opened up to (and including) period $t-1$, rec_{it}^o represents the number of periods since customer i last opened an email. Similarly, freq_{it}^c is computed as the proportion of emails customer i has clicked up to (and including) period $t-1$, and rec_{it}^c measures the number of periods since customer i last clicked an email.

- *RF (covariates)*, in which the opening, clicking and unsubscribing behaviors are modeled as a function of $\mathbf{x}_{it}^u$, $\mathbf{x}_{it}^o$, and $\mathbf{x}_{it}^c$, and the four recency-frequency variables defined above.

- *HMM (static)* in which transitions between states are not allowed (i.e., $\boldsymbol{Q}_{it}$ is the identity matrix). This is simply a finite-mixture model with three latent classes and continuous heterogeneity in the opening and clicking behavior within class.

All four specifications are estimated using a standard MCMC hierarchical Bayes procedure. We tested multiple specifications for unobserved individual heterogeneity by estimating a "full-heterogeneity" model that incorporates a random effect for each of the three behaviors as well as the nested specifications incorporating heterogeneity in opening and clicking only, opening only, or no heterogeneity. When estimating the RF specifications, we find that the fit of the most restricted version (no random effect) is notably better than that of all other specifications.[24]

As with our proposed model, we estimate all the benchmark models using the calibration-period data and assess the prediction for each of the three observed behaviors in the validation period. Note that the RF-based model can only provide one-period-ahead predictions, as the recency and

[23] We do not include a "monetary value" variable because, while the company would receive a higher fee when a customer clicks on more than one deal, such behavior is rare in our data.

[24] We present the best fitting model in the Web Appendix B; all sets of results are available from the authors.

frequency variables summarize behavior up to and including the most recent period. We therefore use actual behavior from the validation period to update these variables.

We compute the mean squared error (MSE) at the individual/period level and then the mean absolute percentage error (MAPE) at the period level for the predictions of each of the three behaviors. We also compute the area under the receiver operating characteristic (ROC) curve (AUC) to assess how well the model identifies customers who are more likely to unsubscribe. The greater the area under the curve, the more accurately the model separates potential churners from those who are more likely to stay. Because we estimate the models in a hierarchical Bayesian framework, we perform this customer-level forecasting exercise for each draw of the Markov chain (after the burn-in period), compute the measures of fit per draw, and report the averages across the draws. Table 9 reports the error measures for all the models, sorted by their accuracy in predicting unsubscribing behavior. It is clear from the table that our model outperforms all other benchmarks in terms of individual-level predictive ability (i.e., it has the lowest MSE across all behaviors and the highest AUC). Our model has the lowest MAPE when predicting unsubscribing behavior, while both RF models have the lowest MAPE when predicting opening and clicking. As judged by AUC, the static model and the RF models do a poor job of identifying which customers more likely to overtly churn.

	MSE			MAPE			AUC
	Open	Click	Unsubs.	Open	Click	Unsubs.	Unsubs.
Proposed	0.087	0.034	0.001	7.9	33.6	75.0	0.720
HMM (static)	0.137	0.045	0.001	32.1	119.2	77.4	0.654
RF (no covariates)	0.114	0.037	0.001	6.1	30.2	80.4	0.544
RF (covariates)	0.114	0.038	0.001	6.7	21.3	77.0	0.544
Lagged covariates	0.189	0.039	0.005	25.4	123.5	2,593.9	0.498

Table 9: Validation-period performance across models, sorted by accuracy in predicting unsubscribing behavior.

These results suggest that, while the RF models do well at predicting the total number of opens and clicks on any given day, the proposed model performs better at identifying *which* customers are more likely to open, click, or unsubscribe (lowest MSE and highest AUC) as well as predicting the total number of unsubscribers. These results are to be expected as the RF models do not separate the two types of churn (overt and silent) and hence cannot accurately identify the antecedents of unsubscribing behavior.

5 Empirical Application 2: Performing Arts Organization

We explore the generalizability of the insights derived from analyzing the daily deals website data by applying our model to data from a very different context. The data comes from a performing arts organization located in New York City. Like most arts organizations, the company keeps the list of all patrons (hereafter customers) who have subscribed to their email service. A customer joins the email list by either buying a ticket (and providing her email address) or by actively subscribing to the newsletter via the company's website. Approximately once a month, the company sends an email to all subscribers. The company generally sends one of two types of emails: (i) "tickets" emails, containing information about upcoming performances and links to the organization's website where tickets can be purchased, and (ii) "information" emails, containing information about performances that have been already announced or about other issues related to the organization. Approximately once a year, the organization sends a "donation" email in which all subscribers are reminded that they can support the organization by donating money.[25] As with the first application, attrition is observed for some customers and is latent for others — customers can unsubscribe from the mailing list or simply ignore emails from the firm.

5.1 Description of the Data

For each email sent by the organization, we observe which customers received the email, whether or not they opened the email and clicked on any content, and whether they unsubscribed from the email service. We have access to twenty months of email activity, ranging from July 2012 to February 2014. In contrast to our first application, we do not observe the entire history of these customers. As a result of a change in email providers, the company does not know when (prior to July 2012) these members joined the email list and has no record of their pre-July 2012 activity. For the purposes of our analysis, we focus on those customers who opened at least one email during the first year of the data and ignore those customers who unsubscribed in the first observation window. We randomly selected 1,000 of these customers to estimate our model. As with our first application, each period corresponds to an instance in which an email was sent to a customer. During our twenty-month observation window, the number of emails received ranged from 2 to 25

[25]The company occasionally sends personalized emails to a small number of "elite" customers; these are emails inviting them to fundraising galas and similar events. Because these emails are rare and targeted at a very specific (and small) set of customers, we ignore these emails and exclude these customers from our analysis.

(which means we have 2 to 25 observations per customer); on average, a customer received 22.2 emails (with a standard deviation of 5.2). During our observation window 66.7% of the emails were "ticket" emails, 22.2% "information" emails, and the remaining 11.1% "donation" emails. The email type is easily inferred by customers (and by us) from the subject line.

Unlike the previous application, customers who unsubscribe from this service can still purchase tickets either from the website or from the ticket office (i.e., customers in this setting do not need to be email subscribers to buy tickets). We collected data from the ticket office on all the ticket purchases by our sample of customers during the period of study and will use this data to validate the (hidden) states identified by our model in Section 5.5.

5.2 Patterns in the data

We observe 22,174 emails sent to our sample of 1,000 customers, of which 25.9% were opened and 3.3% were clicked. (Conditional on being opened, 12.6% of the emails were clicked.) Consistent with our previous application, customers are very heterogeneous in their propensity to open and click, with the percentage of emails opened ranging from 3.7% to 100.0%. More than half of the customers never clicked on an email, and approximately 15% of customers clicked on every email they received. Table 10 shows the summary statistics, across observations and across individuals; these summary statistics are very comparable with those of our first empirical application (Table 1).

	Mean	Std Dev.	Min.	Max	N
Across Observations					
% Open	25.9	43.8	--	--	22,174
% Click	3.3	17.8	--	--	22,174
% Click | Open	12.6	33.2	--	--	7,139
Across Customers					
% Open	27.5	24.0	3.7	100.0	1,000
% Click	3.3	5.9	0.0	40.7	1,000
% Click | Open	21.0	35.3	0.0	100.0	1,000

Table 10: Summary statistics for opening and clicking behavior.

During our observation window, 7.6% of the customers unsubscribed from the service; the average time to unsubscribing was 12 periods after the start of the observation period. We note that the aggregate churn rate does not systematically decrease or increase over time. This is not surprising given that we are dealing with a pool of customers acquired at different points in time.

In a similar manner to Figure 3, we split the sample in terms of whether or not the customer unsubscribed between periods 16 and 25 and explore the dynamics in customer activity in periods 1–15 (Figure 6). Despite the notable differences between these two contexts — daily deals website versus performing arts organization — and despite the fact that customers are observed at different points in their lifetime, the patterns observed in Figure 6 are strikingly similar to those observed for the daily deals website in Figure 3. In this dataset, we once again observe that unsubscribers have a higher probability of opening emails and a lower probability of clicking given opening compared to those customers who do not unsubscribe.

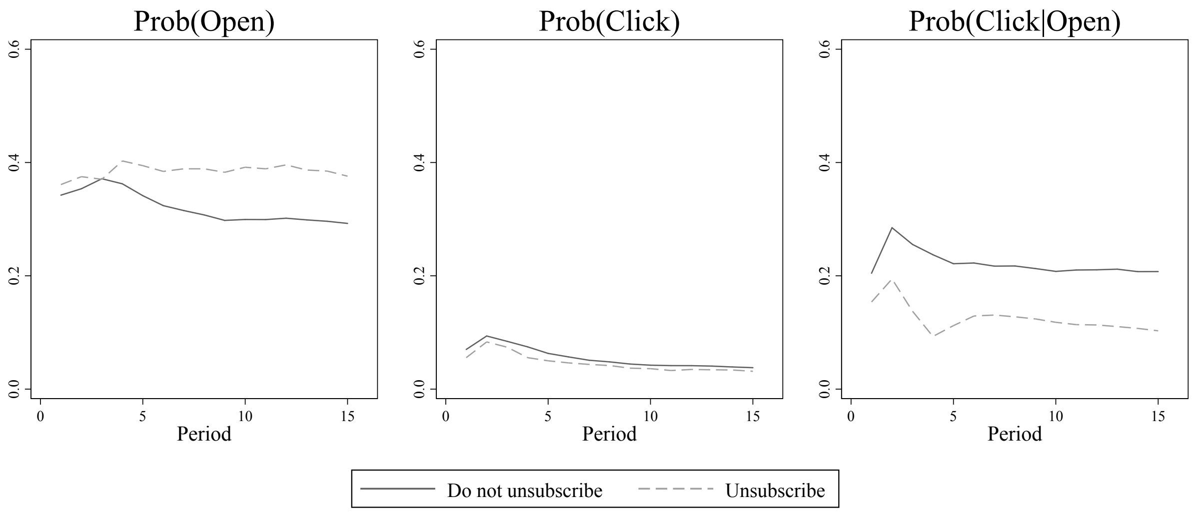

Figure 6: Evolution of the empirical probabilities of opening and clicking for customers, conditional on unsubscribing behavior.

5.3 Model Specification and Selecting the Number of States

In order to account for systematic differences in opening, clicking, and unsubscribing depending on the type of the email, we include email type as a covariate (in the form of dummy variables) affecting all three behaviors as well as the transition matrix. With reference to the notation introduced in Section 3, $\mathbf{x}_{it}^o = \mathbf{x}_{it}^c = \mathbf{x}_{it}^u = \mathbf{x}_{it}^q = [\text{Ticket}, \text{Donation}]$.[26] We estimate the model varying the number of states from one to four, and compute multiple measures of fit.[27] With reference to Table 11,

[26] Information emails serve as the base category.

[27] Because of data limitations (i.e., compared to the first application, there are fewer observations per customer and the clicking rate is lower), it was not possible for us to get reliable estimates when estimating the model with unobserved heterogeneity in all components. As a consequence, we estimate a model with unobserved heterogeneity in the transition probabilities and in opening behavior, but not in clicking and unsubscribing behavior.

we see that the four-state model has the highest log predictive density and the lowest WAIC and MAPE, and that the three-state model has the lowest MSE. For parsimony and ease of comparison across the two applications, we present the results for the three-state model. It is important to note that both the three-state and the four-state model provide very similar insights regarding customer behavior.[28]

	In-sample				Out-of-sample	
# States	Log pred. density	WAIC	MSE	MAPE	MSE	AUC
1	−12,652	25,476	0.204	207.2	0.206	0.554
2	−12,217	24,736	0.153	39.0	0.127	0.717
3	−12,096	24,679	0.144	38.0	0.113	0.672
4	−11,947	24,515	0.145	37.5	0.109	0.674

Table 11: Selecting the number of states.

5.4 Results

In the interest of brevity, we only summarize the key results and discuss the insights relevant to both applications. (See Web Appendix C for a full set of results.) Table 12 presents the posterior means and 95% CPIs for the probabilities of opening, clicking, and unsubscribing in each of the hidden states. The results are consistent with our previous application (Table 4). State 2 captures those customers who seem interested in the product offerings — upon opening an email, they are very likely to click. State 3 captures those customers who have low probabilities of interacting with the company. Customers in state 1 have the highest probability of unsubscribing, the highest probability of opening an email, but the lowest probability of clicking on an email. Looking at the latent state transition probabilities (Table 13), we see that state 3 is very sticky. The probabilities of staying in either of the other two states are also quite high.

Once again we have found three very distinct states of behavior in this customer base: the "engaged" customers (state 2) who actively engage with the service as shown by their high clicking probabilities, the "silently gone" customers (state 3) who do not interact with the firm (either by opening an email or unsubscribing), and the customers who are more "at risk" of terminating their relationship with the company (state 1), with a high probability of opening their emails (higher

[28]See Web Appendix C for a detailed description of the four-state model results and a discussion of its consistency with the three-state model solution.

		Posterior mean	95% CPI	
	State 1	0.623	0.574	0.671
Prob(Open)	State 2	0.251	0.206	0.301
	State 3	0.054	0.040	0.070
	State 1	0.025	0.019	0.031
Prob(Click)	State 2	0.229	0.184	0.278
	State 3	0.004	0.002	0.007
	State 1	0.005	0.003	0.007
Prob(Unsubscribe)	State 2	0.001	0.000	0.003
	State 3	0.002	0.001	0.003

Table 12: Posterior means and 95% CPIs of the state-dependent probabilities of opening, clicking, and unsubscribing.

	To state		
From state	1	2	3
1	0.788	0.010	0.202
	[0.534 , 0.984]	[0.001 , 0.018]	[0.013 , 0.461]
2	0.110	0.767	0.123
	[0.003 , 0.525]	[0.218 , 0.975]	[0.022 , 0.295]
3	0.102	0.009	0.889
	[0.009 , 0.249]	[0.000 , 0.020]	[0.740 , 0.980]

Table 13: Average and 95% heterogeneity interval of the individual posterior means of the transition probabilities.

than those belonging to the two other states), a lower probability of clicking on any of the content of the email than those in the "engaged" group, and the highest risk of unsubscribing from the email service.

5.5 Validation of the States

Unlike the first empirical application, customers of this organization can buy tickets without being subscribed to the email service. Putting it differently, it is possible for customers who appear to be "silently gone" or who have unsubscribed from the email list to keep purchasing tickets from the organization. Thus, churning from the email communication may not mean a complete termination of the relationship with the organization. If that is the case, the implications derived from our analysis could be misleading as these customers might still be of value to the company. To investigate this issue, we collected additional data on individual ticket purchases. We obtained box office data from the beginning of our observation window until three months after the last email was sent, and extracted the records for the customers in our sample. Note that this information was not incorporated in our analysis thus far, and was never used by the company to determine email strategies. In fact, the two datasets reside in different departments within the organization. Accordingly, we can use these purchase data to validate the "silently gone" state and overt churn behavior.

Our main goal is to validate our findings by examining whether the latent states (inferred from changes in opening, clicking, and unsubscribing behaviors) relate to actual purchase behavior. First, we investigate whether customers who overtly churn (unsubscribe from the service) continue purchasing tickets. In other words, we want to confirm that unsubscribing in this setting is a good proxy for true customer churn. We find that, while the customers who unsubscribed during our observation window had purchased tickets prior to unsubscribing, none of these customers ever purchased a ticket after unsubscribing. Second, we analyze actual purchase behavior by customers belonging to each of the hidden states. We use our model predictions to predict the state (either "at risk," "silently gone," and "engaged") a customer belongs to in each period in the calibration sample. We then crosstabulate that information with the number of transactions a customer made during that period. We report in Table 14 the average number of transactions made by the individuals assigned to each state, along with the limits of the 95% interval. We observe that the average number of transactions is significantly higher for customers in the "engaged" state, with "silently gone" customers representing the lowest transaction propensity.

	# Transactions in current period		
State	Average	2.5% percentile	97.5% percentile
"At risk"	0.003	0.002	0.003
"Engaged"	0.010	0.008	0.012
"Silently gone"	0.002	0.001	0.003

Table 14: Number of box office transactions by state.

Thus, our purchase data validates that customers in the "engaged" state are most likely not only to respond to company's communication but also to purchase tickets. Furthermore, we find that customer in the "at risk" rarely buy tickets from the company, despite having the highest rate of email opening. Customers who have churned overtly from the email communication also did so in terms of purchases. Because we did not use the ticket purchasing data when calibrating the model and identifying the customers' latent states, we believe this analysis shows that the states identified by our model using email engagement data reflects customer purchase behavior in the form of monetary transactions.

6 Discussion

In this paper we have examined customer behavior in hybrid contractual/noncontractual settings where the loss of some customers is observed by the firm while the loss of others is unobserved. To accommodate both types of attrition we propose a hidden Markov model (HMM) that allows us to separate those customers who are "engaged" with the firm from those who are "at risk" of churning overtly and those who are "silently gone." The model captures customer dynamics in the level of engagement and in the risk of leaving the company by allowing customers to transition among latent states. Moreover, the model measures the impact of factors such as the quality of the firm's communications on the individual transitions between states, as well as on customer behavior given state membership.

We apply the model to two different business settings — a daily deals website and a performing arts organization. While these two contexts promote and sell very different types of goods and services, target different sets of customers, and communicate at very different time intervals, they both share the commonality that customers can choose to either stop interacting with the organization by unsubscribing from the email list (i.e., overtly churning) or by simply ignoring incoming emails from the organization (i.e., silently churning). In the first application, the email communication is the product itself and, thus, termination of the communications implies the end of any relationship with the firm. In the second application, unsubscribing implies termination of the email-based communications with the firm, but need not imply an end to event attendance. We validate our latent states by demonstrating that both overt and silent churn from the email communication also imply churn from monetary transactions.

In both empirical applications, we consistently find that the focal company's customer base can be characterized by three latent groups of customers, which we label "engaged," "silently gone," and those who are "at risk" of termination. In contrast to previous studies that have not separated the two types of churn, we find that in hybrid settings a high level of activity — captured in our applications by actively opening the emails received from the firm — is not necessarily a good indicator of future profitability as it is associated with a higher risk of overtly churning.

Treating these settings as noncontractual (i.e., ignoring overt churn) will miss the opportunity to understand, predict, and manage this churn behavior. On the other hand, treating them as contractual settings (i.e., ignoring latent attrition) will mix together active and "silently gone" customers and miss the opportunity to manage each group appropriately. Moreover, by mixing these two types of customers, it becomes difficult to identify the factors/behaviors that predict overt churn because inactive customer will be seeing as "stayers."

Using characteristics of the email content as covariates in both the state-dependent probabilities and in the state transition probabilities, we explore the antecedents of the two types of churn. We find that when engaged customers receive emails with better content, they are less likely to transition to the "silently gone" state. More customized emails can help re-engage customers who are at risk of churning. Using a set of counterfactual analyses, we demonstrate that the firm could manage its customer base by transitioning customers to states with a lower risk of churn (overt or latent) and increase profitability from clicks.

In addition to providing superior predictions of customer behavior, our model offers several managerially relevant insights. We are able to identify those customers who are at risk of formally terminating their relationship with the firm (unsubscribers) and determine which of the remaining customers are silent churners. Because each type of churner raises different flags in terms of their behaviors, disentangling the two types of churn is crucial when guiding companies in their retention efforts. Across both applications, we highlight the importance of measuring and leveraging multiple dimensions of behavior. For example, if companies only focus on metrics that directly link to revenue — that is, focusing on clicking and ignoring opening behavior — their insights into customer churn would be misleading. For example, analyzing clicking behavior alone does not allow us to disentangle unsubscribers from "stayers" (Figures 3 and 6). Moreover, understanding the dynamics in opening behavior is also crucial for detecting those customers who are "silently gone."

This research is particularly relevant for managers involved in customer relationship management. We not only offer insights into customer behavior but also provide a tool for quantifying and predicting customer activity that can identify those who might be at risk of actively leaving the company and detecting those customers who should no longer be considered as part of the firm's active customer base. As we demonstrate, improving the product offering (i.e., the email content) might reduce the risk of unsubscribing. However, our results suggest that it is likely that such efforts would have little impact on a large portion of the customer base who are "silently gone" and are not expected to open any email in the future (and are thus unlikely to be exposed to the improved content). As a consequence, firms ought to think of different ways to re-engage those customers classified as "silently gone." One approach would be to alter the header of emails, signaling "silently gone" customers that the offerings have changed. Firms may also want to consider varying the frequency with which emails are sent or contact these customers through other channels. While our data are not rich enough to measure the impact of such marketing efforts, our proposed model could easily be used to measure their effects.

The managerial implications for email marketing are also noteworthy. Identifying "silently gone" customers is increasingly important for email marketing firms, who are interested in removing "dead" customers from their mailing lists so as to maximize the overall deliverability of their email communications by keeping their "sender ranking" at high levels (see, for example, Return Path 2014). Mailbox providers use recipients' engagement (email opening and clicking), or lack thereof, to determine whether to re-route the sender's emails to spam folders. Accordingly, identifying "silently gone" customers and removing them from the firm's email recipient list is crucial to the proper management of the list.

Our research builds on the existing literature on modeling customer attrition (e.g., Ascarza et al. (2016), Braun and Schweidel (2011), and Neslin et al. (2006) for observed attrition, and Fader et al. (2010), and Schweidel et al. (2014) for latent attrition). We generalize existing methods by proposing a modeling framework that can accommodate both types of attrition. From a methodological perspective, it complements the growing literature in marketing that highlights the usefulness of HMMs for understanding customer dynamics in a variety of marketing problems (e.g., Ascarza and Hardie 2013, Montoya et al. 2010, Netzer et al. 2008, Schwartz et al. 2014, Schweidel et al. 2011, Zhang et al. 2014). More broadly, our work also relates to the literature on email marketing (e.g., Bonfrer and Drèze 2009, Drèze and Bonfrer 2008, Kumar et al. 2014) as email communications and newsletters are one example of hybrid settings.

With the popularity of free or freemium services, the number of hybrid settings in the marketplace is rapidly growing. In addition to the applications investigated in this research, there are many other business settings where the loss of some customers is observed by the firm while the loss of others is unobserved. Examples include social networking sites (e.g., Facebook, LinkedIn, Tinder), digital services (e.g., Dropbox, Gmail), online games, and the majority of apps. The customer activities observed in these contexts are not necessarily opening and clicking on emails, but rather logins, likes, review posts, friends requests, rounds played, in-app purchases, and so on. How would our modeling framework separate both types of churn when behaviors are different? First, incorporating these activities in our modeling framework is straightforward, as the HMM structure can be easily adapted to include as many state-dependent behaviors as needed. Second, our model would separate the two types of churn provided the observed behaviors (or a combination of them) enable us to discriminate between overt churners and engaged customers. For example, consider the case of a a social network, one in which overt churn would be the closing of an account. Observed behaviors in this case would take the form of logins, searches, friends requests, and in-app messages. Our modeling framework would easily identify "silent churners" by very low levels of any activity. Customers "at risk" may be characterized by frequent logins and searches, but low levels of social activity in terms of building and increasing their social network by inviting friends and accepting friend requests. We encourage future work that applies our modeling framework to these settings, as doing so would help managers better understand which behaviors should be tracked and leveraged.

While our model allows us to separate overt from silent churners, our data and modeling efforts shed little light on why some customers decide to churn overtly and others silently. We postulate that one of the reasons for churning overtly rather than silently could be the degree to which the customer is negatively affected by the amount and/or the content of the communication with the firm. For example, an offensive communication or total mismatch between the firm's communication and the customer preferences is more likely to lead to overt rather than a silent churn. Because our data does not allow us to directly investigate this issue, we leave this for future research.

The current research provides a first step in exploring an emerging type of business setting, one in which customers can decide whether to overtly or silently churn from the firm. Such settings break the traditionally defined distinction between contractual and noncontructual settings. We encourage future research to continue investigating these interesting settings and identify additional drivers that help distinguish between these two types of churners.

References

Ascarza E (2016) Retention futility: Targeting high risk customers might be ineffective. `http://ssrn.com/abstract=2759170` (Accessed 2017-02-22).

Ascarza E, Hardie BGS (2013) A joint model of usage and churn in contractual settings. *Marketing Sci.* 32(4):570–590.

Ascarza E, Iyengar R, Schleicher M (2016) The perils of proactive churn prevention using plan recommendations: Evidence from a field experiment. *J. Marketing Res.* 53(1):46–60.

Blattberg RC, Getz G, Thomas JS (2001) *Customer Equity: Building and Managing Relationships as Valuable Assets* (Harvard Business School Press, Boston, MA).

Blattberg RC, Kim B-D, Neslin SA (2008) *Database Marketing: Analyzing and Managing Customers* (Springer, New York, NY).

Bonfrer A, Drèze X (2009) Real-time evaluation of e-mail campaign performance. *Marketing Sci.* 28(2):251–263.

Borle S, Singh SS, Jain DC (2008) Customer lifetime value measurement. *Management Sci.* 54(1):100–112.

Braun M, Schweidel DA (2011) Modeling customer lifetimes with multiple causes of churn. *Marketing Sci.* 30(5): 881–902.

Braun M, Schweidel DA, Stein E (2015) Transaction attributes and customer valuation. *J. Marketing Res.* 52(6):848–864.

Buckinx W, Verstraeten G, Van den Poel D (2007) Predicting customer loyalty using the internal transactional database. *Expert Systems with Applications* 32(1):125–134.

Chen K, Hu YH, Hsieh YC (2015) Predicting customer churn from valuable B2B customers in the logistics industry: A case study. *Inform. Systems and e-Business Management* 13(3):475–494.

Coussement K, De Bock KW (2013) Customer churn prediction in the online gambling industry: The beneficial effect of ensemble learning. *J. Bus. Res.* 66(9):1629–1636.

Coussement K, Van den Poel D (2008) Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications* 34(1):313–327.

Danaher PJ, Bonfrer A, Dhar S (2008) The effect of competitive advertising interference on sales for packaged goods. *J. Marketing Res.* 45(2):211–225.

Deming WE, Glasser GJ (1968) A Markovian analysis of the life of newspaper subscriptions. *Management Sci.* 14(6):B283–B293.

Drèze X, Bonfrer A (2008) An empirical investigation of the impact of communication timing on customer equity. *J. Interactive Marketing.* 22(1):36–50.

Dwyer, F. R. (1989). Customer lifetime valuation to support marketing decision making. *J. of Direct Marketing.* 3(4):8–15.

Fader PS, Hardie BGS (2007) How to project customer retention. *J. Interactive Marketing.* 21(1):76–90.

Fader PS, Hardie BGS (2010) Customer-base valuation in a contractual setting: The perils of ignoring heterogeneity. *Marketing Sci.* 29(1):85–93.

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.* 29(6):1086–1108.

Fader PS, Hardie BGS, Sen S (2014) Stochastic models of buyer behavior. Winer RS, Neslin, SA, eds. *The History of Marketing Science.* (World Scientific Publishing, Singapore), 165–205.

Gelman A, Carlin JB, Stern HA, Dunson DB, Vehtari A, Rubin DB (2013) *Bayesian Data Analysis* (3rd edn) (Chapman & Hall/CRC, London).

Gupta S, Lehmann DR (2005) *Managing Customers as Investments* (Wharton School Publishing, Upper Saddle River, NJ).

Knox G, Van Oest R (2014) Customer complaints and recovery effectiveness: A customer base approach. *J. Marketing* 78(5):42–57.

Kumar V, Zhang X, Luo A (2014) Modeling customer opt-in and opt-out in a permission-based marketing context. *J. Marketing Res.* 51(4):403–419.

Lemmens A, Croux C (2006) Bagging and boosting classification trees to predict churn. *J. Marketing Res.* 43(2):276–286.

Mason CH (2003) Tuscan Lifestyles: Assessing customer lifetime value. *J. Interactive Marketing* 17(4):54–60.

Montoya R, Netzer O, Jedidi K (2010) Dynamic allocation of pharmaceutical detailing and sampling for long-term profitability. *Marketing Sci.* 29(5):909–924.

Neslin SA, Gupta S, Kamakura W, Lu J, Mason CH (2006) Defection detection: Measuring and understanding the predictive accuracy of customer churn models. *J. Marketing Res.* 43(2):204–211.

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

Pfeifer PE, Carraway RL (2000). Modeling customer relationships as Markov chains. *J. Interactive Marketing* 14(2): 43–55.

Reinartz WJ (1999) *Customer Lifetime Value Analysis: An Integrated Empirical Framework for Measurement and Explanation* (Unpublished Ph.D dissertation, College of Business Administration, University of Houston).

Return Path (2014) *2014 Sender Score Benchmark Report.* `http://returnpath.com/research/email-marketing-sender-score-benchmark-report-2014/` (Accessed 2015-05-13).

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

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

Schweidel DA, Fader PS, Bradlow ET (2008) Understanding service retention within and across cohorts using limited information. *J. Marketing* 72(1):82–94.

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

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

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

Silverpop (2013) *2013 Email Marketing Metrics Benchmark Study: An Analysis of Messages Sent Q1–Q4, 2012.* `http://www.silverpop.com/marketing-resources/white-papers/all/2013/2013-EM-Benchmark/` (Accessed 2015-05-13)

Zhang JZ, Netzer O, Ansari A (2014) Dynamic targeted pricing in B2B relationships. *Marketing Sci.* 33(3):317–337.

Zucchini W, MacDonald IL (2009) *Hidden Markov Models for Time Series: An Introduction Using R.* (CRC Press, Boca Raton, FL).

Web Appendix A: Model Estimation

In this appendix we describe the hierarchical Bayesian framework used to estimate the model parameters.

So as to ensure identification of the states (i.e., prevent label switching), we restrict the opening probabilities to be decreasing in the relationship states.[A1] We therefore reparameterize $\boldsymbol{\zeta^o}$ in the following manner,

$$\zeta_k^o = \begin{cases} \zeta'^o_1 & \text{if } k = 1 \\ \zeta'^o_{k-1} - \exp\left(\zeta'^o_k\right) & \text{for } k = 2, \ldots, K\,, \end{cases}$$

and estimate $\boldsymbol{\zeta'^o}$ instead. Note that this restriction does not impose any a priori assumption on the model.

Let $\boldsymbol{\Omega}$ denote the following (population-level) parameters: $\boldsymbol{\zeta'^o}$ (opening), $\boldsymbol{\zeta^c}$ (clicking given open), $\boldsymbol{\zeta^u}$ (unsubscribing given open), $\boldsymbol{\phi}$ (transition probabilities), $\boldsymbol{\rho}$ (initial probabilities), $\boldsymbol{\beta}$ (effect of covariates on the state-dependent probabilities), and $\boldsymbol{\delta}$ (effect of covariates on the transition probabilities). The vector $\boldsymbol{\xi}_i = \{\boldsymbol{\eta}_i, \boldsymbol{\psi}_i\}$ contains all the individual-level unobserved parameters for customer i. Let $\boldsymbol{\xi} = \{\boldsymbol{\xi}_i\}_{i=1,\ldots,I}$ denote all the individual-level parameters (where I is the number of customers in the calibration sample), and $\boldsymbol{\Sigma}_\xi$ denote the variance-covariance matrix for cross-sectional heterogeneity. The full joint posterior distribution can be written as

$$f(\boldsymbol{\Omega}, \boldsymbol{\xi}|\text{data}) \propto \left\{ \prod_{i=1}^{I} \mathscr{L}_i(\boldsymbol{\Omega}, \boldsymbol{\xi}_i \,|\, \text{data}) f(\boldsymbol{\xi}_i \,|\, \boldsymbol{\Sigma}_\xi) \right\} f(\boldsymbol{\Sigma}_\xi) f(\boldsymbol{\Omega})\,,$$

where $\mathscr{L}_i(\boldsymbol{\Omega}, \boldsymbol{\xi}_i \,|\, \text{data})$ is defined in (11). The term $f(\boldsymbol{\xi}_i \,|\, \boldsymbol{\Sigma}_\xi)$ denotes the prior (or mixing) distribution for $\boldsymbol{\xi}_i$, which is assumed to follow a multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Sigma}_\xi$. The terms $f(\boldsymbol{\Omega})$ and $f(\boldsymbol{\Sigma}_\xi)$ denote the (hyper)priors for the population parameters. Uninformative (vague) priors are used for all parameters. We assume $\boldsymbol{\Sigma}_\xi$ has an inverse-Wishart prior with degrees of freedom $\text{df} = K(K-1) + 7$ and scale matrix $\boldsymbol{R}$, with $\text{diag}(\boldsymbol{R}) = \mathbb{1}_{K(K-1)+2}/(3 \times \text{df})$, where $\mathbb{1}_A$ denotes a $1 \times A$ vector of ones. We assume that $\boldsymbol{\Omega}$ has a multivariate normal prior with mean $\boldsymbol{\mu}_\Omega$ and variance-covariance matrix $\boldsymbol{\Sigma}_\Omega$. The values of $\boldsymbol{\mu}_\Omega$ and $\boldsymbol{\Sigma}_\Omega$ were chosen to ensure uninformative priors in the transformed space. The mean is specified as $\boldsymbol{\mu}_\Omega = \left[0 \times \mathbb{1}_K\,, 0.1 \times \mathbb{1}_K\,, -3 \times \mathbb{1}_K\,, 0.5 \times \mathbb{1}_{K(K-1)}\,, -10 \times \mathbb{1}_{K-1}\,, -1 \times \mathbb{1}_{C^\delta}\,, 0.1 \times \mathbb{1}_{C^o+C^c+C^u}\right]$,

[A1] By ensuring ordering in one of the state-dependent behaviors (e.g., probability of opening given state membership) we prevent label switching without imposing any restriction on the relationships among behaviors.

where C^δ is the number of covariates included in the transition probability equation multiplied by $K(K-1)$, and C^u, C^o, and C^c are the number of covariates incorporated in the equations for each of the three observed behaviors (excluding the intercept), multiplied by K. The variance-covariance matrix is specified as $\mathrm{diag}(\boldsymbol{\Sigma}_\Omega) = \left[1.2 \times \mathbb{1}_K\,,\, 0.3 \times \mathbb{1}_K\,,\, 0.5 \times \mathbb{1}_K\,,\, 0.01 \times \mathbb{1}_{K(K-1)}\,,\, 0.1 \times \mathbb{1}_{K-1}\,,\, 0.1 \times \mathbb{1}_{C^\delta}\,,\, 0.3 \times \mathbb{1}_{C^o+C^c+C^u}\right]$.

Since there are no closed-form expressions for the posterior distributions of $\boldsymbol{\xi}$ and $\boldsymbol{\Omega}$, we use a Gaussian random-walk Metropolis-Hasting algorithm to draw from these distributions. We draw recursively from the following posterior distributions:

- [Metropolis-Hastings]
$$f(\boldsymbol{\Omega} \,|\, \boldsymbol{\mu}_\Omega, \boldsymbol{\Sigma}_\Omega, \boldsymbol{\xi}, \mathrm{data}) \propto \exp\left(.5(\boldsymbol{\Omega} - \boldsymbol{\mu}_\Omega)' \boldsymbol{\Sigma}_\Omega^{-1} (\boldsymbol{\Omega} - \boldsymbol{\mu}_\Omega)\right) \prod_{i=1}^{I} \mathscr{L}_i(\boldsymbol{\Omega}, \boldsymbol{\xi}_i \,|\, \mathrm{data}).$$

- [Metropolis-Hastings]
$$f(\boldsymbol{\xi}_i \,|\, \boldsymbol{\Sigma}_\xi, \boldsymbol{\Omega}, \mathrm{data}) \propto \exp\left(.5\boldsymbol{\xi}_i' \boldsymbol{\Sigma}_\xi^{-1} \boldsymbol{\xi}_i\right) \mathscr{L}_i(\boldsymbol{\Omega}, \boldsymbol{\xi}_i \,|\, \mathrm{data}), \;\forall\, i.$$

- [Gibbs]
$$f(\Sigma_\xi \,|\, \boldsymbol{\xi}, \boldsymbol{R}, \mathrm{df}) \sim \text{inv-Wishart}\left(\sum_{i=1}^{I} \boldsymbol{\xi}_i' \boldsymbol{\xi}_i + \mathrm{df}\boldsymbol{R}^{-1}, \mathrm{df} + I\right).$$

For the Metropolis-Hastings steps, we follow the procedure proposed by Atchadé (2006) and adapt the tuning parameters in each iteration to get an acceptance rate of approximately 20%. In the empirical analyses reported in the paper, we ran the simulation for 500,000 iterations. The first 400,000 iterations were used as a "burn-in" period, and the last 100,000 iterations were used to estimate the conditional posterior distributions.

References

Atchadé YF (2006) An adaptive version for the Metropolis adjusted Langevin algorithm with a truncated drift. *Methodol. Comput. Appl. Probab.* 8(2):235–254.

Web Appendix B: Additional Results: Empirical Application 1

B1 Additional results: Three-state specification

For ease of exposition, the main document shows the impact of changes in the covariates on the transition probabilities. Table B1 reports the posterior estimates of the coefficients for the covariates included in the state transition process of the proposed model with three states.

Variable	Posterior mean	95% CPI	
Lag(QualStock)			
From state 1 to 1	−0.842	−2.232	0.567
From state 1 to 2	**1.456**	0.210	2.827
From state 2 to 1	0.975	−0.360	2.264
From state 2 to 2	0.768	−0.849	2.237
From state 3 to 1	**−1.589**	−2.701	−0.387
From state 3 to 2	0.777	−0.815	2.304
Number of periods since last email			
From state 1 to 1	−0.038	−0.710	1.005
From state 1 to 2	−1.130	−2.639	0.638
From state 2 to 1	0.572	−0.276	1.584
From state 2 to 2	−0.437	−1.198	0.294
From state 3 to 1	0.449	−0.073	1.019
From state 3 to 2	0.114	−1.056	1.080

Table B1: Posterior means of the coefficients for the covariates included in the state transition process. Numbers in bold are associated with 95% CPIs not including 0.

B2 Results for the one-, two-, and four-state specifications

In this appendix we present the results for the one-, two-, and four-state specifications of our proposed model. We focus on the two main sets of results: the state-specific probabilities for all three behaviors (open, click, and unsubscribe) and the transition probabilities among the latent states.

B2.1 Model with one state ($K = 1$)

We start by describing the results of the most restricted model, in which only one latent state is allowed. As such, it is a static specification and there are no parameters governing dynamics.

		Posterior mean	95% CPI	
Prob(Open)	State 1	0.286	0.268	0.304
Prob(Click)	State 1	0.059	0.048	0.072
Prob(Unsubscribe)	State 1	0.002	0.001	0.002

Table B2: State-dependent probabilities for the one-state model.

B2.2 Model with two states ($K = 2$)

We now turn to the results of the simplest dynamic specification, one with two hidden states. Combining the results from Tables B3 and B4, and comparing them with those obtained when three hidden states are allowed (Tables 4 and 5 in the main manuscript) we observe that a state of high activity is identified (state 1). This state capture customers with high activity in *all three behaviors*. Unlike the insights derived from the three-state specification, the model with two states fails to separate the very engaged customers from those who open frequently, but rarely click and are more likely to unsubscribe. The model clearly captures the state of the "silently gone" customers, who rarely interact with the service. As we find in the three-state specification, such a state is highly absorbing.

		Posterior mean	95% CPI	
Prob(Open)	State 1	0.600	0.553	0.641
	State 2	0.068	0.053	0.084
Prob(Click)	State 1	0.178	0.135	0.229
	State 2	0.026	0.011	0.043
Prob(Unsubscribe)	State 1	0.003	0.001	0.005
	State 2	0.001	0.000	0.001

Table B3: State-dependent probabilities for the two-state model.

From state	To state 1	2
1	0.796	0.204
	[0.546 , 0.987]	[0.013 , 0.454]
2	0.040	0.960
	[0.001 , 0.141]	[0.859 , 0.999]

Table B4: Mean transition probabilities and the 95% heterogeneity interval of individual posterior means for the two-state model.

B2.3 Model with four states ($K = 4$)

We finally consider the results of an HMM with four states and compare them with the more parsimonious specification of an HMM with three states. Comparing Tables B5 and B6 with Tables 4 and 5, we see that allowing additional flexibility in the model (i.e., adding an additional state) sees the "silently gone" state in the three-state model divided into two states, both of which are very sticky and have low probabilities of any kind of activity. Moreover, in contrast to the three-state specification, the four-state model is not as effective at separating the "at risk" and "engaged" states, as suggested by the distributions of clicking and unsubscribing probabilities— they have closer posterior means and wider 95% CPIs than those obtained with the three-state model.

		Posterior mean	95% CPI	
Prob(Open)	State 1	0.656	0.617	0.695
	State 2	0.683	0.637	0.729
	State 3	0.069	0.055	0.084
	State 4	0.080	0.067	0.093
Prob(Click)	State 1	0.115	0.068	0.171
	State 2	0.239	0.140	0.344
	State 3	0.015	0.007	0.031
	State 4	0.001	0.000	0.002
Prob(Unsubscribe)	State 1	0.003	0.002	0.005
	State 2	0.004	0.001	0.007
	State 3	0.001	0.000	0.002
	State 4	0.001	0.000	0.002

Table B5: State-dependent probabilities for the four-state model.

From state	To state 1	2	3	4
1	0.568	0.205	0.041	0.187
	[0.128 , 0.992]	[0.000 , 0.581]	[0.000 , 0.080]	[0.007 , 0.303]
2	0.186	0.370	0.265	0.179
	[0.102 , 0.304]	[0.225 , 0.588]	[0.057 , 0.453]	[0.087 , 0.229]
3	0.005	0.061	0.903	0.031
	[0.000 , 0.013]	[0.001 , 0.206]	[0.715 , 0.996]	[0.003 , 0.067]
4	0.039	0.006	0.013	0.941
	[0.022 , 0.060]	[0.004 , 0.009]	[0.008 , 0.02]	[0.917 , 0.964]

Table B6: Mean transition probabilities and the 95% heterogeneity interval of individual posterior means for the four-state model.

B3 Benchmark model details and results

In Section 4.6 we compare the out-of-sample predictions generated by our model with benchmark models in which each of the three behaviors — opening, clicking, and unsubscribing — are modeled as a function of lagged covariates and/or past behavior (recency and frequency). We model each of the three behaviors — unsubscribing, opening, and clicking — using binary logit models. More formally,

$$\bar{o}_{it} = P(Y^o_{it} = 1) = \frac{e^{\mathbf{z}^o_{it}\boldsymbol{\lambda}^o_i}}{1 + e^{\mathbf{z}^o_{it}\boldsymbol{\lambda}^o_i}} \tag{B1}$$

$$\bar{c}_{it} = P(Y^c_{it} = 1) = \begin{cases} \dfrac{e^{\mathbf{z}^c_{it}\boldsymbol{\lambda}^c_i}}{1 + e^{\mathbf{z}^c_{it}\boldsymbol{\lambda}^c_i}} & \text{if } y^o_{it} = 1 \\ 0 & \text{if } y^o_{it} = 0\,, \end{cases} \tag{B2}$$

$$\bar{u}_{it} = P(Y^u_{it} = 1) = \begin{cases} \dfrac{e^{\mathbf{z}^u_{it}\boldsymbol{\lambda}^u_i}}{1 + e^{\mathbf{z}^u_{it}\boldsymbol{\lambda}^u_i}} & \text{if } y^o_{it} = 1 \\ 0 & \text{if } y^o_{it} = 0\,, \end{cases} \tag{B3}$$

where $\mathbf{z}^o_{it}$, $\mathbf{z}^c_{it}$, and $\mathbf{z}^u_{it}$ represent the covariates affecting each of the behaviors. Using the same logic as in (8), the probability that customer i has observed behavior $\boldsymbol{y}_{it} = [\,y^o_{it}, y^c_{it}, y^u_{it}\,]$ in period t is

$$\begin{aligned} P(\boldsymbol{Y}_{it} = \boldsymbol{y}_{it}) = \mathbb{1}(y^o_{it} = 1)\bar{o}_{it}\Big\{&\big[\mathbb{1}(y^c_{it} = 1)\bar{c}_{it} + \mathbb{1}(y^c_{it} = 0)(1 - \bar{c}_{it})\big] \\ &\times \big[\mathbb{1}(y^u_{it} = 1)\bar{u}_{it} + \mathbb{1}(y^u_{it} = 0)(1 - \bar{u}_{it})\big]\Big\} \\ &+ \mathbb{1}(y^o_{it} = 0)(1 - \bar{o}_{it})\,. \end{aligned} \tag{B4}$$

It follows that customer i's likelihood function is

$$\mathscr{L}_i(\boldsymbol{\lambda_i} \,|\, \text{data}) = \prod_{t=1}^{T_i} P(\boldsymbol{Y}_{it} = \boldsymbol{y}_{it}), \tag{B5}$$

where $\boldsymbol{\lambda_i}$ contains $\boldsymbol{\lambda}_i^u$, $\boldsymbol{\lambda}_i^o$, and $\boldsymbol{\lambda}_i^c$ (i.e., the parameters in (B1)–(B3)). As described in Section 4.6, we consider three specifications of this model: lagged covariates, RF (no covariates) and RF (covariates). Tables B7–B9 report the parameter estimates for these models.

Our fourth benchmark model, HMM (static) is a version of our HMM in which transitions between states are not allowed (i.e., $\boldsymbol{Q}_{it}$ is the identity matrix). The associated parameter estimates are reported in Table B10.

Behavior	Variable	Posterior mean	95% CPI	
Open	Intercept	**−0.895**	−0.921	−0.867
	Sunday	**0.103**	0.037	0.167
	Lag(QualStock)	**−1.437**	−1.626	−1.283
	Number of periods since last email	**0.089**	0.035	0.143
	Lag(Sunday)	0.030	−0.026	0.085
	Lag2(QualStock)	**−0.071**	−0.130	−0.010
	Lag(Number of periods since last email)	**0.072**	0.011	0.131
Click | Open	Intercept	**−0.907**	−1.121	−0.672
	Sunday	**−0.194**	−0.335	−0.042
	log(#deals)	**0.495**	0.331	0.665
	Discount	**−0.056**	−0.084	−0.033
	Time left	**0.002**	0.000	0.005
	Food	**−0.074**	−0.105	−0.046
	Fitness	**−0.039**	−0.062	−0.020
	Source	−0.001	−0.015	0.017
	Order	**−0.030**	−0.047	−0.016
	Lag(QualStock)	**3.052**	1.848	4.215
	Number of periods since last email	0.008	−0.081	0.095
	Lag(Sunday)	**−0.131**	−0.262	−0.003
	Lag(log(#deals))	0.025	−0.084	0.139
	Lag(Avg. Discount)	0.168	−0.635	0.935
	Lag(Avg. Time left)	−0.010	−0.026	0.008
	Lag(%Food)	−0.062	−0.265	0.139
	Lag(%Fitness)	**1.079**	0.738	1.421
	Lag(Avg. Source)	0.098	−0.088	0.301
	Lag2(QualStock)	−0.005	−0.114	0.107
	Lag(Number of periods since last email)	0.006	−0.107	0.115
Unsubscribe | Open	Intercept	**−5.149**	−5.546	−4.773
	Avg. Discount	**1.248**	0.499	1.951
	Avg. Time left	0.005	−0.062	0.067
	%Food	**1.244**	0.702	1.819
	%Fitness	**0.935**	0.318	1.567
	Avg. Source	0.164	−0.384	0.738
	Sunday	**−0.603**	−1.182	−0.003
	log(#deals)	−0.284	−0.654	0.093
	Lag(QualStock)	**1.915**	0.870	2.943
	Number of periods since last email	0.161	−0.171	0.446
	Lag(Avg. Discount)	0.723	−0.034	1.383
	Lag(Avg. Time left)	−0.058	−0.165	0.031
	Lag(%Food)	**−0.661**	−1.190	−0.092
	Lag(%Fitness)	**0.980**	0.147	1.756
	Lag(Avg. Source)	0.005	−0.647	0.707
	Lag(Sunday)	0.151	−0.356	0.650
	Lag(log(#deals))	**0.703**	0.222	1.209
	Lag2(QualStock)	0.089	−0.274	0.541
	Lag(Number of periods since last email)	−0.090	−0.543	0.273

Table B7: Parameter estimates for the lagged covariates model. Numbers in bold are associated with 95% CPIs not including 0.

Behavior	Variable	Posterior mean	95% CPI	
Open	Intercept	$\mathbf{-1.980}$	-2.064	-1.894
	rec^{o}	$\mathbf{-0.072}$	-0.083	-0.062
	freq^{o}	$\mathbf{3.512}$	3.392	3.633
	rec^{c}	0.004	-0.001	0.009
	freq^{c}	$\mathbf{-1.106}$	-1.275	-0.939
Click | Open	Intercept	$\mathbf{-0.426}$	-0.587	-0.266
	rec^{o}	$\mathbf{0.052}$	0.030	0.075
	freq^{o}	$\mathbf{-1.733}$	-1.944	-1.520
	rec^{c}	$\mathbf{-0.083}$	-0.096	-0.071
	freq^{c}	$\mathbf{2.535}$	2.269	2.804
Unsubscribe | Open	Intercept	$\mathbf{-4.694}$	-5.270	-4.143
	rec^{o}	$\mathbf{0.064}$	0.003	0.122
	freq^{0}	-0.636	-1.345	0.064
	rec^{c}	-0.002	-0.041	0.034
	freq^{c}	0.539	-0.458	1.512

Table B8: Parameter estimates for the RF (no covariates) model. Numbers in bold are associated with 95% CPIs not including 0.

Behavior	Variable	Posterior mean	95% CPI	
Open	Intercept	**−1.923**	−2.010	−1.835
	rec^o	**−0.078**	−0.089	−0.068
	freq^o	**3.435**	3.320	3.550
	rec^c	0.002	−0.003	0.007
	freq^c	**−0.909**	−1.085	−0.731
	Sunday	**0.131**	0.056	0.205
Click | Open	Intercept	−0.008	−0.248	0.236
	rec^o	**0.052**	0.030	0.073
	freq^o	**−1.751**	−1.951	−1.558
	rec^c	**−0.082**	−0.094	−0.070
	freq^c	**2.625**	2.324	2.907
	Sunday	**−0.200**	−0.333	−0.062
	log(#deals)	**0.409**	0.262	0.562
	Discount	**−0.019**	−0.036	−0.004
	Time left	**−0.020**	−0.035	−0.005
	Food	0.001	−0.001	0.004
	Fitness	**−0.050**	−0.083	−0.020
	Source	**−0.024**	−0.046	−0.007
	Order	0.000	−0.003	0.002
Unsubscribe | Open	Intercept	**−4.826**	−5.426	−4.212
	rec^o	**0.064**	0.001	0.123
	freq^o	**−0.697**	−1.391	−0.015
	rec^c	−0.001	−0.040	0.035
	freq^c	0.720	−0.340	1.719
	Avg. Discount	0.525	−0.653	1.725
	Avg. Time left	−0.006	−0.078	0.056
	%Food	0.474	−0.131	1.057
	%Fitness	−0.144	−1.205	0.926
	Avg. Source	−0.080	−0.727	0.631
	Sunday	**−0.620**	−1.316	−0.022
	log(#deals)	−0.084	−0.392	0.240

Table B9: Parameter estimates for the RF (covariates) model. Numbers in bold are associated with 95% CPIs not including 0.

	Segment 1	Segment 2	Segment 3
Effect on Opening			
Sunday	**−1.307**	**0.145**	**2.522**
	[−2.449 , −0.195]	[0.046 , 0.252]	[1.403 , 3.520]
Effect on Clicking, given Opening			
Sunday	0.604	**−0.228**	1.270
	[−0.76 , 2.02]	[−0.414 , −0.062]	[−0.429 , 2.901]
log(#deals)	0.100	**0.740**	0.826
	[−1.498 , 1.858]	[0.498 , 0.982]	[−0.777 , 2.414]
Discount	0.112	**0.147**	−0.255
	[−0.16 , 0.506]	[0.041 , 0.264]	[−1.291 , 0.413]
Time left	−0.082	**0.003**	−0.166
	[−0.218 , 0.020]	[0.000 , 0.006]	[−0.561 , 0.010]
Food	0.135	**−0.125**	−0.113
	[−0.102 , 0.506]	[−0.163 , −0.087]	[−0.855 , 0.511]
Fitness	0.022	**0.148**	0.412
	[−0.243 , 0.337]	[0.057 , 0.242]	[−0.216 , 1.660]
Source	0.025	0.003	−0.092
	[−0.223 , 0.303]	[−0.035 , 0.041]	[−0.956 , 0.577]
Order	**−0.160**	**−0.056**	−0.392
	[−0.347 , −0.011]	[−0.086 , −0.028]	[−1.22 , 0.012]
Effect on Unsubscribing, given Opening			
Avg. Discount	0.057	−0.404	−0.012
	[−1.726 , 1.803]	[−1.898 , 1.224]	[−1.903 , 1.975]
Avg. Time left	0.081	−0.023	−0.339
	[−0.173 , 0.366]	[−0.127 , 0.063]	[−1.134 , 0.308]
%Food	−0.334	0.765	−0.108
	[−2.096 , 1.497]	[−0.155 , 1.696]	[−1.605 , 1.449]
%Fitness	1.222	**−2.197**	−1.377
	[−0.562 , 3.14]	[−3.904 , −0.566]	[−3.529 , 0.867]
Avg. Source	−0.702	0.068	−0.365
	[−2.568 , 1.277]	[−1.021 , 1.290]	[−2.141 , 1.566]
Sunday	−0.216	**−1.033**	−0.103
	[−1.931 , 1.48]	[−2.096 , −0.221]	[−1.692 , 1.402]
log(#deals)	−1.273	0.530	0.501
	[−2.379 , 0.017]	[−0.259 , 1.605]	[−1.687 , 2.609]

Table B10: Parameter estimates (posterior means) for the HMM (static) model. Numbers in bold are associated with 95% CPIs (in brackets) not including 0.

Web Appendix C: Additional Results: Empirical Application 2

C1 Additional results: Three state specification

In this appendix we report the results from the second empirical application that were not discussed in the main manuscript, namely the initial state probabilities (Table C1) and the posterior estimates of the coefficients for the covariates included in the models of state-dependent behavior (Table C2) for the proposed model with three states.

	Posterior mean	95% CPI	
State 1	0.486	0.428	0.547
State 2	0.291	0.236	0.356
State 3	0.223	0.151	0.295

Table C1: Initial state probabilities.

	State		
	"At Risk"	"Engaged"	"Silently gone"
Effect on Opening			
Ticket	-0.086	$\mathbf{0.852}$	$\mathbf{1.261}$
	[-0.231 , 0.064]	[0.470 , 1.256]	[0.731 , 1.941]
Donation	-0.070	-0.543	-0.951
	[-0.274 , 0.146]	[-1.295 , 0.101]	[-2.361 , 0.152]
Effect on Clicking, given Opening			
Ticket	0.331	0.357	$\mathbf{-1.623}$
	[-0.160 , 0.862]	[-0.883 , 1.575]	[-2.808 , -0.377]
Donation	$\mathbf{-1.794}$	$\mathbf{-2.277}$	-1.452
	[-3.021 , -0.754]	[-3.604 , -1.004]	[-3.763 , 1.090]
Effect on Unsubscribing, given Opening			
Ticket	-0.052	-1.569	0.483
	[-0.770 , 0.647]	[-3.621 , 0.292]	[-0.923 , 1.900]
Donation	0.724	0.937	0.450
	[-0.004 , 1.478]	[-1.572 , 3.120]	[-2.097 , 2.688]

Table C2: Posterior means of the effect of the covariates on the state-dependent probabilities. Numbers in bold are associated with 95% CPIs (in brackets) not including 0.

C2 Results for the four-state specification

Recall from Section 5 that the models with three and four latent states provided very similar measures of fit. While the model with three states had the best fit in terms of MSE, the model with four states had lower log predictive density and WAIC. In this section we present the results of the model with four states and discuss how they differ from those obtained using the three-state model.

		Posterior mean	95% CPI	
Prob(Open)	State 1	0.637	0.589	0.682
	State 2	0.992	0.981	0.998
	State 3	0.092	0.069	0.118
	State 4	0.013	0.007	0.019
Prob(Click)	State 1	0.017	0.011	0.023
	State 2	0.897	0.814	0.963
	State 3	0.904	0.820	0.973
	State 4	0.218	0.183	0.256
Prob(Unsubscribe)	State 1	0.007	0.007	0.005
	State 2	0.009	0.007	0.001
	State 3	0.000	0.000	0.000
	State 4	0.001	0.001	0.001

Table C3: State-dependent probabilities.

Comparing Table C3 with Table 12, we see that allowing additional flexibility in the model (i.e., adding an additional state) results in the "engaged" state from the three-state model being divided into two states, one state in which customers open almost every e-mail (state 2) and another state in which customers open less frequently (state 3). Moreover, if we label state 1 as "at risk" and state 4 as "silently gone," these two states clearly coincide in terms of customer behavior with their corresponding states in the three-state model. The transition matrix (Table C4) is consistent with this intuition. The transition probabilities related to the "at risk" and "silently gone" states are consistent with those obtained in the three-state model. Regarding the two "engaged" states, the state with very high opening behavior is not very sticky, while the other state is more stable.

	To state			
From state	1	2	3	4
1	0.965	0.000	0.000	0.035
	[0.867 , 0.998]	[0.000 , 0.000]	[0.000 , 0.001]	[0.002 , 0.132]
2	0.000	0.319	0.170	0.511
	[0.000 , 0.000]	[0.149 , 0.411]	[0.037 , 0.36]	[0.242 , 0.800]
3	0.155	0.034	0.809	0.002
	[0.013 , 0.395]	[0.001 , 0.181]	[0.541 , 0.983]	[0.001 , 0.003]
4	0.000	0.001	0.002	0.996
	[0.000 , 0.000]	[0.000 , 0.003]	[0.001 , 0.005]	[0.993 , 0.998]

Table C4: Mean transition probabilities and the 95% heterogeneity interval of individual posterior means.