MARKETING SCIENCE
Vol. 27, No. 5, September–October 2008, pp. 829–843
ISSN 0732-2399 | EISSN 1526-548X | 08 | 2705 | 0829

DOI 10.1287/mksc.1070.0328

A Bivariate Timing Model of Customer Acquisition and Retention

David A. Schweidel
Department of Marketing, University of Wisconsin–Madison School of Business, Madison, Wisconsin 53706, dschweidel@bus.wisc.edu

Peter S. Fader, Eric T. Bradlow
Department of Marketing, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104 {faderp@wharton.upenn.edu, ebradlow@wharton.upenn.edu}

Two widely recognized components, central to the calculation of customer value, are acquisition and retention propensities. However, while extant research has incorporated such components into different types of models, limited work has investigated the kinds of associations that may exist between them. In this research, we focus on the relationship between a prospective customer's time until acquisition of a particular service and the subsequent duration for which he retains it, and examine the implications of this relationship on the value of prospects and customers.

To accomplish these tasks, we use a bivariate timing model to capture the relationship between acquisition and retention. Using a split-hazard model, we link the acquisition and retention processes in two distinct yet complementary ways. First, we use the Sarmonov family of bivariate distributions to allow for correlations in the observed acquisition and retention times *within a customer*; next, we allow for differences *across* customers using latent classes for the parameters that govern the two processes. We then demonstrate how the proposed methodology can be used to calculate the discounted expected value of a subscription based on the time of acquisition, and discuss possible applications of the modeling framework to problems such as customer targeting and resource allocation.

Key words: customer acquisition; customer retention; customer relationship management; stochastic models
History: This paper was received May 2, 2006, and was with the authors 7 months for 3 revisions; processed by Alan Montgomery. Published online in *Articles in Advance* April 7, 2008.

1. Introduction

Service acquisition and retention have been closely tied to key managerial metrics such as the value of the customer base (e.g., Gupta and Zeithaml 2006, Gupta et al. 2004) and critical managerial decisions such as resource allocation (e.g., Blattberg and Deighton 1996). These two constructs—the time that elapses before a prospective customer acquires a particular service and the subsequent duration for which a customer retains service before dropping it—may be related. Despite the obvious appeal (and likelihood) of such a relationship, few published papers model the interplay between these two behaviors (Jain and Singh 2002); instead, most treat acquisition and retention probabilities as independent (e.g., Gupta et al. 2004, Blattberg and Deighton 1996). This assumption, however, may not accurately reflect the true behavioral propensities of customers, thereby adversely affecting the firm's forecasts and subsequent marketing decisions. That is, it is not just an academic question of bias, but a managerial one of potentially significant impact.

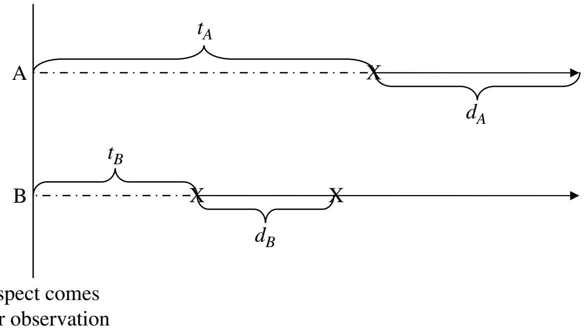
In this research, we develop a joint timing model to explore the acquisition and retention of prospective customers for a contractual service provider. Specifically, we consider multiple aspects of customer behavior in the acquisition and retention of a service. First, we allow for customers' *binary decision* to ever acquire service. Among those customers who acquire service, we allow for *duration dependence* in both the acquisition and retention processes. That is, the likelihood that a prospect acquires a particular service may change as he goes even longer without that service. Similarly, after service acquisition, the likelihood of discarding the service may also depend on how long the customer has subscribed to that service. Next, we incorporate the notion of *correlated processes*—that is, the correlation that may exist between the observed acquisition and retention durations, as well as the parameters that govern them. Last, we account for *unobserved heterogeneity*, as customers may have different propensities for acquiring and discarding a particular service.

While unobserved heterogeneity and duration dependence have received much attention in extant timing models (e.g., Morrison and Schmittlein 1980), limited research has focused on their role in the acquisition and retention processes or the relationship between them. To illustrate the importance of jointly considering acquisition and retention, consider two hypothetical prospects, A and B, for a contractual service provider, depicted in Figure 1.

The times that elapsed between the beginning of acquisition efforts for prospects A and B and when they acquired service are denoted by t_A and t_B, respectively, and the durations for which they maintained service are given by d_A and d_B, respectively. Based on these observations, should the firm pursue those prospects who have gone without the service for longer periods of time? If there were no relationship between the time until acquisition and length of service retention, the company would expect both prospects to retain service for the same length of time (if they acquire the service at all). However, it may be that (as depicted in Figure 1, where $d_A > d_B$) prospects like A, who are slow to acquire service, maintain it for a longer period of time (e.g., "cautious but loyal"), whereas prospects who acquire service quickly (like prospect B) also discard it quickly (e.g., "hit and run").

We can broadly classify potential customer-level relationships between the acquisition and retention outcomes into three types. First, a *negative* relationship may exist between time until acquisition and retention duration. As such, customers who were slow to acquire a service would not be expected to retain service for as long as a customer who acquired it earlier. If true, the firm may want to devote resources toward acquiring younger prospects rather than older prospects based on their expected tenure. On the other hand, in the case of a *positive* relationship (as reflected in Figure 1), customers who acquired a particular service later will have a longer expected tenure (ET). In this scenario, older prospects may be tempting targets. Even though they have not yet acquired service, if they do, they will be expected to retain service longer and therefore may generate greater long-term revenue than early acquirers. Finally, there may be no relationship between acquisition and retention duration, as is commonly assumed. As such, the time at which prospective customers acquire a service is uninformative of their tenure and targeting based on the age of the prospect plays no significant role. Providing a modeling framework within which to understand the relationship between the time until acquisition and length of service retention is the primary objective of this research.

Figure 1 **Depiction of Acquisition-Retention Relationship**



*Acquisition times are denoted by t_A and t_B, retention durations by d_A and d_B.

In exploring the relationship between acquisition of service and duration of retention, we consider two possible sources of correlation that can exist. First, for a given customer, the time of acquisition and retention durations may be related to each other. To allow for this correlation, we make use of the Sarmanov family of multivariate distributions (Kotz et al. 2000, Lee 1996), which has begun to make its way into the marketing literature (Danaher 2006, Danaher and Hardie 2005, Park and Fader 2004). To allow for unobserved heterogeneity across customers in terms of their acquisition and retention propensities, we use latent classes (e.g., Kamakura and Russell 1989). Thus, in addition to the correlation between these processes *within* customers (using the Sarmanov family), the latent classes allow for a separate correlation between the acquisition and retention processes *across* customers.

In considering the time until acquisition and subsequent duration of service (as in Figure 1), the need to account for censoring is highlighted. While many prospective customers may not acquire a particular service during the observation period, they might acquire it after the observation period ends (i.e., right censoring of the acquisition process). Of those customers who acquire service, some may retain it through the end of the observation period but drop service at a later (unobserved) time (i.e., right censoring of the retention process). As a result, standard exploratory bivariate analyses, such as computing the correlation between observed acquisition and retention times, may require that we discard a significant portion of the observed data (i.e., those prospects who do not acquire service as well as those who acquire service and maintain it through the observation period). Our joint stochastic model does not suffer this limitation and provides a natural way to account for censored observations via the inclusion of survival functions for both processes in the likelihood.

The remainder of this paper proceeds as follows. In §2, we review existing literature related to the acquisition and retention of services, as well as other applications of bivariate timing models. Section 3 discusses the data used in our empirical analyses. In §4, we present the development of the proposed joint timing model. The empirical analyses are detailed in §5, including a discussion of a series of models (nested and otherwise) that are estimated. In §6, we demonstrate how our proposed model can be incorporated into managerial decisions. Managerial implications and directions for future research are discussed in §7.

2. Previous Research

Despite the importance of considering both acquisition and retention as part of a comprehensive customer valuation model, few studies have taken the (possible) relationship between them into account. Those that do deal with both processes often assume complete independence across them. For instance, Gupta et al. (2004) propose a rich framework for determining the value of an entire customer base, but they assume that the acquisition and retention processes are independent. Blattberg and Deighton (1996) discuss the need for firms to balance their marketing expenditures between acquiring new customers and retaining existing customers. They propose a framework to achieve an appropriate balance across acquisition and retention spending, but they do not consider any links that may exist between the two processes. Berger and Nasr-Bechwati (2001) apply a similar approach to the problem of resource allocation but, like Blattberg and Deighton (1996), they do not shed light on the underlying relationship between acquisition and retention that may exist at the customer level.

Perhaps the first paper to consider separate models for each process is Hansotia and Wang (1997), who discuss the importance of acquiring those customers who will be most profitable on the basis of their lifetime value. Their treatment of lifetime value, however, considers the acquisition process as binary (yes/no) and independent of the retention process, which is modeled using a right-censored Tobit model.

The most significant contribution in this area is from Thomas (2001), who proposes a methodology for linking the customer acquisition and retention processes that is closest in spirit to our model, as it incorporates heterogeneity and correlation between the acquisition and retention processes; yet, it is still substantially different from our approach. Thomas (2001) uses a Tobit model with selection to jointly model the acquisition and retention of an optional membership available to individuals who already belong to an organization. This model is tantamount to using a binary (yes/no) probit model for the customer's decision to acquire membership and a Tobit model with right-censoring to model the time for which he retains it, where the errors for the probit and Tobit models are correlated. Latent classes are used to incorporate heterogeneity into the acquisition and retention processes.

Reinartz et al. (2005) extend Thomas' (2001) earlier model to simultaneously model acquisition, retention, and customer profitability. Actions taken by the firm and customer, as well as customer characteristics, are assumed to affect the acquisition process, the retention process, and customer profitability. The three outcome variables (acquisition, duration, and profitability) are assumed to have a correlated error structure. The authors explore how the level of investment and resource allocation between acquisition and retention can differentially affect customer acquisition, retention, and profitability. As does Berger and Nasr-Bechwati (2001), this research emphasizes the need to jointly model acquisition and retention for marketing decisions.

While Thomas (2001) and Reinartz et al. (2005) account for the link between acquisition and retention (through correlated errors), there are several major limitations to the generalizability of their approach. First, acquisition is solely considered as a binary variable. Consequently, only the decision to acquire service (yes/no) is modeled rather than the *time* at which acquisition occurred. As such, these models assume that retention duration and profitability do not depend on the *time* of acquisition. There are, however, several situations including service acquisition (which is the context of our empirical application), in which the time that has elapsed since a customer came under observation and when it actually acquired the particular service may be known. Other ways in which customers may come under the observation of firms are via rented mailing lists (e.g., Bitran and Mondschein 1996) or through opt-in activities (e.g., Milne and Rohm 2000). The length of this acquisition period may convey information about subsequent activity, such as retention (in a contractual setting) or purchasing (in a transactional setting), which can offer insights about the value of prospective customers.

These models also do not facilitate incorporation of duration dependence or time-varying covariates into the acquisition/retention process. Prospective customers may change in their likelihood of acquiring (or discarding) service over time. In addition, the company may use promotional activities to entice customers to acquire a particular service, the effect of which may accumulate over time. This will also affect a customer's likelihood of retaining service while the promotion is active. In this research, we use the proportional hazards framework (e.g., Seetharaman and Chintagunta 2003), which easily accommodates both duration dependence (via the baseline hazard function) and time-varying covariates, to construct the marginal acquisition and retention processes and then allow for correlation between the processes.

We are not the first to develop a bivariate timing model, nor the first to use the Sarmanov family of distributions. Chintagunta and Haldar (1998) study the relationship in purchase timing of two related categories. Using the Farlie-Gumbel-Morgenstern family of bivariate distributions, they were the first in marketing to use a bivariate hazard function to understand the relationship between two timing processes. It should be noted, however, that the Farlie-Gumbel-Morgenstern bivariate distributions do not generally yield marginal distributions that match the univariate densities used to construct the joint distribution (Park and Fader 2004). Given the prevalence of right censoring in the acquisition and subsequent retention processes, marginal distributions that are the same as the specified univariate distributions (or are least available in closed form) are highly desirable, and they are provided naturally (and by design) by the Sarmanov family of bivariate distributions (Kotz et al. 2000, Lee 1996).

Previous marketing applications of the Sarmanov family of distributions have focused on noncontractual settings (Danaher 2006, Danaher and Hardie 2005, Park and Fader 2004). Park and Fader (2004) develop a bivariate timing model in which they examine intervisit times at multiple Web sites and find the need to allow for correlation across sites. Danaher and Hardie (2005) use the Sarmanov family of distributions to develop a bivariate counting (not timing) model to examine copurchasing behavior. Danaher (2006) generalizes the Sarmanov distribution beyond the bivariate case and presents a multivariate counting model to predict the audience for online advertising campaigns. One limitation of these current marketing applications of the Sarmanov family of distributions (to date) has been the lack of covariates in either process. We overcome this limitation by using the proportional hazards framework to modify the marginal processes.

Although the bivariate modeling approaches discussed above are similar to the present research, the fundamental research problems differ. This distinction is evident when comparing the timing of the two processes. While previous applications examined behaviors that occurred *contemporaneously*, in this research we are examining the relationship between two *sequential* behaviors: acquisition and then retention. Because of this sequential nature, customers cannot discard service until after it has been acquired. Additionally, there is a possibility that customers will never acquire service and therefore will not face either an acquisition timing or retention timing decision. We explicitly allow for this by using a split-hazard model (e.g., Sinha and Chandrashekaran 1992) for the acquisition process, coupled with a proportional hazard model for the retention process, ultimately developing a bivariate correlated split-hazard model that is appropriate for sequential timing processes.

In summary, the current research builds on previous literature to develop a joint timing model of service acquisition and retention. As previous research has demonstrated the importance of considering both acquisition and retention simultaneously, we build a customer-level model that allows us to explore the possible underlying relationship between these two processes. Understanding this link will allow firms to determine if prospective customers are worth the resources needed to convert them to active customers. In the following sections, we explicitly describe our model and demonstrate how the proposed modeling framework can be incorporated into managerial decisions.

3. Data

Monthly subscription data were provided by a major provider of telecommunications services. From the subscription records, we constructed a calibration data set by sampling half of the customers (who began service with the firm between February 2002 and September 2002) and used the remaining half for holdout analysis. These customers were observed from the time at which they started service through May 2004. To demonstrate the proposed methodology, we examine those customers who did not initially subscribe to Home Box Office (hereafter HBO, a cable network service to which customers may optionally subscribe for additional premium programming), but came under observation as an HBO prospect by subscribing to other services (e.g., high-speed Internet or digital cable). There were 6,211 customers in the calibration sample and 6,236 customers in our holdout sample.

There are two data features that we do not explore in this research but which are worth noting. First, our sample does not consider those customers who began their tenure as subscribers of HBO. The behavior of these customers may systematically differ from those customers who subscribe to HBO after becoming customers. Second, there are a limited number of customers (approximately 2% of the sample) who discard and resubscribe to service, with an average duration of 6.4 months between subscription spells. In our model, we restrict attention to the initial acquisition and retention spells. Thus, our results may not accurately reflect the behavior of all customers, but only the initial subscription to HBO of those customers who did not sign up for HBO service at the beginning of their tenure as a customer. The modeling framework, however, can be extended to examine these groups in more detail and is an area for future research.

Time until acquisition was computed by counting the number of months that elapsed between the customer's first month of (non-HBO) service with the provider and when the customer, if ever, added HBO. Note, as before, that time until acquisition may be right censored, as some customers will not acquire the service during the observation period. In a similar vein, the duration for which a customer retains HBO service was computed by counting the number of months for which a customer keeps service following its acquisition. Retention times may also be right censored.

In Figures 2(a) and 2(b), we provide an overview of the acquisition and retention processes, respectively, by examining the empirical hazard rates. That is, Figure 2(a) shows the observed probability of a customer acquiring service each month, given that they had not yet acquired service. Figure 2(b) shows the observed probability of customers discarding service each month, conditional on maintaining it through the previous month.

Immediately, we see that the empirical hazard rates for both acquisition and retention are not constant. For service acquisition, at first glance, it appears that customers become less likely to acquire service over time. For retention, we note that the churn rate is initially low before increasing after the third month of service. This is consistent with marketing activity run by the service provider, which offered a price promotion for the month of acquisition and the subsequent two months of service. After this promotional period, the empirical hazard rate decreases, indicating that customers are less likely to discard service the longer that they have had it. Note that we must exercise caution in drawing conclusions from these plots, as our observations of duration dependence could also be attributable to the sparse nature of the data and unobserved heterogeneity, particularly with regard to retention. It may be, for example, that customers with a high propensity for dropping service do so much earlier than customers with a low propensity for doing so, rather than customers becoming less likely to discontinue service over time (e.g., Follman and Goldberg 1988). Both of these factors provide a strong motivation for a formal model of both processes.

Although these two aggregate plots shed light on the marginal univariate processes, they provide no information about the relationship between acquisition and retention. In the calibration sample, 84.1% of prospects did not acquire service by the end of the observation period. In Table 1, we provide an overview of the acquisition and retention behavior by the 15.9% of the prospects in the calibration sample who did acquire service.

Three particular data features are worth noting. First, the majority of customers do not acquire service during the observation period. As Thomas (2001)

Figure 2(a) Empirical Hazard Rate for Service Acquisition

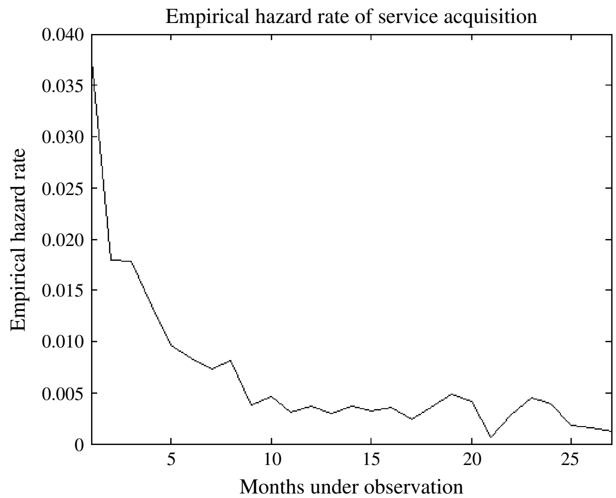


Figure 2(b) Empirical Hazard Rate for Service Retention

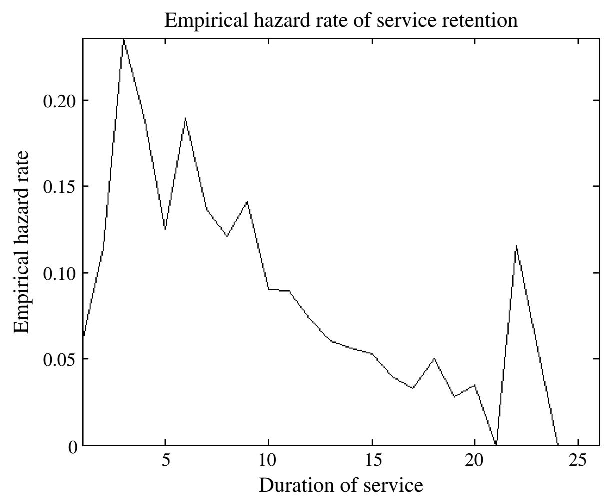


Table 1 Overview of Acquisition and Retention Timing for Prospective Customers Who Came Under Observation from February 2002 Through September 2002

	Retention (%)			
Acquisition	Within 3 months of acquisition	Between 4 and 12 months of acquisition	After 12 months of acquisition	Right censored
Between 1 and 3 months	19.3	17.7	3.6	4.5
Between 4 and 12 months	9.6	16.5	1.5	5.7
After 12 months	6.3	4.5	0	10.8

discusses, analyses that ignore censored data will therefore yield misleading results. Second, of those customers who do acquire HBO service, most acquire it early on. Taken together with Figure 2(a), these observations seem to indicate that negative duration dependence is present in the acquisition process. Last, the retention behavior appears to vary between those prospects who acquired service early and those who acquired it later. Of those prospects who acquired service within 3 months of coming under observation, 43% discarded service within the first 3 months, in contrast to the 29% of customers who discarded service after acquiring it between 4 and 12 months after coming under observation. For the prospects who acquired service before the end of the observation period, using these discrete cells, it appears that acquisition and retention behavior are not independent ($\chi^2 = 175.1$, $p < 0.001$). Together, this seems to indicate a positive relationship between time of acquisition and duration of retention. However, as previously stated, this could be attributable to unobserved heterogeneity or correlation between the two processes. We now turn to the development of our proposed timing model, which allows us to address this question and its managerial importance.

4. Model Development

In this section, we build a customer-level continuous-time joint model for the acquisition and retention of a service that incorporates duration dependence and correlation in the two processes, which is then adapted to discrete-time (e.g., monthly) data. We then show how we incorporate unobserved cross-sectional heterogeneity via latent classes.[1]

4.1. A Customer-Level Bivariate Model of Acquisition and Retention Behavior

Our model considers the acquisition and retention of a particular service for prospective customers who have come under observation. We first present the bivariate timing model for a household who eventually acquires service, and then, in §4.3, discuss how we account for those who never acquire service. We begin by assuming that customers who acquire service do so according to a parametric distribution, where the probability of acquiring service at time t_A is given by

$$f(t_A \mid \Theta) = S_A(t_A - 1 \mid \Theta) - S_A(t_A \mid \Theta) \quad \text{for } t_A = 1, 2, \ldots, T, \tag{1}$$

where $S_A(t_A \mid \Theta)$ is the survival function of the parametric distribution and T is the length of the observation period (e.g., number of months). Once customers acquire service, they discard it according to a different parametric distribution, which has a baseline hazard function denoted $h_R(t_R \mid \Phi)$. To incorporate time-varying activity (or covariates in general) such as the three-month introductory price promotion previously mentioned into the retention process, we use a proportional hazards regression using a baseline hazard function, $h_R(t_R \mid \Phi)$. This results in the survival function for the retention process

$$S_R(t_R \mid \beta, \Phi, \mathbf{X}(t)) = e^{-\sum_{v=1}^{t_R} e^{\beta \mathbf{X}(v)} (\int_{v-1}^{v} h_R(u \mid \Phi)\, du)}, \tag{2}$$

where β is the impact of the time-varying marketing activities, denoted $\mathbf{X}(t)$. In our empirical application, $\mathbf{X}(t)$ contains the introductory promotional activity, where $\mathbf{X}(t) = 1$ for $t = 1, 2$ and 0 otherwise. If $\beta > 0$, customers are more likely to discard service, whereas they are less likely to do so if $\beta < 0$. The duration of service retention d_h is therefore distributed as

$$\begin{aligned} g(t_R \mid \beta, \Phi, \mathbf{X}(t)) &= S_R(t_R - 1 \mid \beta, \Phi, \mathbf{X}(t)) - S_R(t_R \mid \beta, \Phi, \mathbf{X}(t)) \\ &\quad \text{for } t_R = 1, 2, \ldots, T. \end{aligned} \tag{3}$$

Because the promotional activity was continually offered throughout our data period, its impact on the acquisition process cannot be estimated. Had the service provider not offered the same introductory promotion regardless of time, the same approach could have been used to model the effect of time-varying covariates on the acquisition process as well (i.e., extending Equation (1) to Equation (2)). Within our modeling framework, we could then decompose the effect of promotional activity into two parts: the impact of promotions prior to acquisition (e.g., "How much more likely are prospects to acquire service under the promotion?") and subsequent to service acquisition (e.g., "How much less likely are customers to drop service during the promotion?"), a fascinating and important question. Unfortunately, while our framework allows for this kind of analysis, our current data does not.

4.2. Incorporating Correlation Between the Acquisition and Retention Processes

The model presented in Equations (1)–(3) outlines a flexible bivariate timing model that allows for time-varying covariates. However, it treats time until acquisition and retention duration as independent processes. To model the possible correlation between the observed acquisition and retention outcomes, we use the Sarmanov family of bivariate distributions (e.g., Lee 1996, Park and Fader 2004). The Sarmanov family works by defining

$$f(x, y) = f_x(x) \times f_y(y) \times \{1 + \omega \phi_x(x) \phi_y(y)\}, \tag{4}$$

[1] For ease of notation, we suppress the subscript of the latent class while developing the customer-level model.

where $f_x(x)$ and $f_y(y)$ are univariate probability density functions, and $\phi_x(x)$ and $\phi_y(y)$ are bounded mixing functions such that $\int_{-\infty}^{\infty} f_z(z)\phi_z(z)\,dz = 0$ for $z = x, y$. In order for $f(x, y)$ to be a bivariate density function, $\phi_x(x)$, $\phi_y(y)$, and ω must satisfy the condition $1 + \omega\phi_x(x)\phi_y(y) \geq 0$ for all values of x and y. As such, ω can be interpreted as the unnormalized correlation, or covariance, between $f_x(x)$ and $f_y(y)$, our desired goal.

While there are other families of bivariate distributions generated from the combination of two univariate distributions (e.g., Farlie 1960; Johnson and Kotz 1975, 1977), the main benefit that the Sarmanov family of distributions provides for our applications is that the marginal distributions are guaranteed to take on the same forms of the corresponding univariate densities, as can be seen by the construction in Equation (4). This is the only class of multivariate distributions that offers this property, which greatly aids in the closed-form computation of censored events, a feature of computational importance. Consider a customer who does not acquire service during the observation period. This customer may acquire service after the observation period and then proceed to retain service according to the retention process. To calculate the probability that the customer acquires service after T_h, we must integrate over all possible unobserved retention durations. By using a Sarmanov bivariate distribution, this marginal likelihood will be equal to the likelihood computed from the univariate acquisition distribution, which is directly available.

Lee (1996) demonstrates how to find the mixture functions for different distributions $f_x(x)$ and $f_y(y)$. Lee shows that the mixing distribution for any univariate density function $f(x)$ is given by

$$\phi(x) = f(x) - \int_{-\infty}^{\infty} f^2(t)\,dt. \tag{5}$$

Replacing the integral in Equation (5) with a summation provides the general form of the mixing function for a discrete distribution $f(x)$. Then, replacing $f(x)$ in Equation (5) with the probability mass functions given in Equations (1) and (3) yields the mixing functions for our acquisition and retention processes, denoted $\phi_A(t_A)$ and $\phi_R(t_R)$, respectively[2]

$$\begin{aligned}\phi_A(t_A \mid \Theta) &= S_A(t_A - 1 \mid \Theta) - S_A(t_A) \\ &\quad + 2\sum_{i=1}^{\infty}[S_A(i \mid \Theta)(S_A(i-1 \mid \Theta) - S_A(i \mid \Theta))] - 1 \end{aligned}\tag{6}$$

[2] A detailed derivation of the mixing functions can be found in the appendix.

and

$$\begin{aligned}&\phi_R(t_R \mid \beta, \Phi, \mathbf{X}(t)) \\ &\quad = S_R(t_R - 1 \mid \beta, \Phi, \mathbf{X}(t)) - S_R(t_R \mid \beta, \Phi, \mathbf{X}(t)) \\ &\qquad + 2\sum_{i=1}^{\infty}[S_R(i \mid \beta, \Phi, \mathbf{X}(t))(S_R(i-1 \mid \beta, \Phi, \mathbf{X}(t)) \\ &\qquad\qquad - S_R(i \mid \beta, \Phi, \mathbf{X}(t)))] - 1. \end{aligned}\tag{7}$$

The joint distribution of acquisition time and duration of service at the customer level is then given by

$$\begin{aligned}&j(t_A, t_R \mid \omega, \Theta, \beta, \Phi, \mathbf{X}(t)) \\ &\quad = f(t_A \mid \Theta) g(t_R \mid \beta, \Phi, \mathbf{X}(t)) \\ &\qquad \times \{1 + \omega\phi_A(t_A \mid \Theta)\phi_R(t_R \mid \beta, \Phi, \mathbf{X}(t))\}. \end{aligned}\tag{8}$$

4.3. Accounting for the Subscription Decision and Censored Observations

The customer-level joint timing model presented in Equation (8) assumes that all customers will *eventually* acquire service. While some customers may acquire service (i.e., subscribe to HBO, as in our example) after the observation period, some prospective customers may *never* acquire service. To allow for this possibility, we adopt a split-hazard model in which we assume that customers acquire service with a probability z (e.g., Kamakura et al. 2004, Sinha and Chandrashekaran 1992). Hence, prospective customers will never acquire service with a probability of $1 - z$.

Let t_h denote the length of time that has elapsed between when customer h comes under observation and when the customer acquired service, and its accompanying acquisition censoring variable c_h^A set equal to 0 if the observation is not censored. If the customer does *not* acquire service while under observation for a length of T_h, let $t_h = T_h$ and let $c_h^A = 1$. For these customers, there are two possibilities: Either they will never acquire service, or they will do so at some point in the interval $t_h \in (T_h + 1, \infty)$. If it is the latter, they may retain service for any duration d_h. The customer-level likelihood is therefore given by

$$\begin{aligned}&p_1(t_h, d_h \mid z, \omega, \Theta, \beta, \Phi, \mathbf{X}(t)) \\ &\quad = (1 - z) + z\sum_{t_A = T_h + 1}^{\infty}\sum_{t_R = 1}^{\infty} j(t_A, t_R \mid \omega, \Theta, \beta, \Phi, \mathbf{X}(t)) \\ &\quad = (1 - z) + z \cdot S_A(T_h \mid \Theta). \end{aligned}\tag{9}$$

The first term $1 - z$ accounts for the probability that a customer never acquires service, while the second term allows for the probability that they acquire service after the observation period. As we use a Sarmanov bivariate distribution, the latter component comes from the marginal acquisition distribution and is given by the survival probability of the acquisition distribution.

The time at which a customer discontinues service is given by $t_h + d_h$. If the customer retains service throughout the observation period, let $d_h = T_h - t_h$ and the retention censoring variable $c_h^R = 1$; otherwise, let $c_h^R = 0$. The probability of observing a customer who acquires service at time t_h and maintains it through the observation period ($c_h^A = 0$, $c_h^R = 1$) is given by

$$p_2(t_h, d_h \mid z, \omega, \Theta, \beta, \Phi, \mathbf{X}(t)) = z \cdot \sum_{t_R = d_h + 1}^{\infty} j(t_h, t_R \mid \omega, \Theta, \beta, \Phi, \mathbf{X}(t)). \tag{10}$$

Last, there are customers who acquire the service and subsequently discard it during the observation period ($c_h^A = 0$, $c_h^R = 0$). In this case, neither the acquisition nor the retention process is right censored. Thus, the probability that customer h starts service at t_h and maintains service for a duration of d_h is

$$p_3(t_h, d_h \mid z, \omega, \Theta, \beta, \Phi, \mathbf{X}(t)) = z \cdot j(t_h, d_h \mid \omega, \Theta, \beta, \Phi, \mathbf{X}(t)). \tag{11}$$

Combining Equations (9)–(11), the customer-level likelihood can then be written as

$$P(z, \omega, \Theta, \beta, \Phi, \mathbf{X}(t) \mid t_h, d_h, c_h^A, c_h^R, T_h, \mathbf{X}(t)) = \begin{cases} p_1(t_h, d_h), & \text{if } c_h^A = 1 \\ p_2(t_h, d_h), & \text{if } c_h^A = 0 \text{ and } c_h^R = 1 \\ p_3(t_h, d_h), & \text{if } c_h^A = 0 \text{ and } c_h^R = 0. \end{cases} \tag{12}$$

4.4. Allowing for Unobserved Heterogeneity and "Double Correlation"

The bivariate timing model outlined in Equations (1)–(12) relaxes the assumption of independence between the two processes and allows for the within-customer correlation in acquisition and retention. This, however, may not be the only type of relationship that exists between the two processes. In addition to a correlation in the observed durations (as detailed in Equations (5)–(8)), there may also be correlation in the latent acquisition and retention propensities. Park and Fader (2004) refer to the correlation in the underlying parameters as *linked propensities*.

To account for differences *across* customers, we use a latent class model with s unobserved classes (e.g., Thomas 2001, Kamakura and Russell 1989). Therefore, in addition to allowing for unobserved heterogeneity in customers' propensities to acquire and discard service, as the parameters governing both the acquisition and retention processes are segment specific, the latent classes also allow us to capture the linked propensities. For example, one latent class may be characterized by high acquisition likelihood and high retention likelihood, whereas another may be marked by high acquisition likelihood and low retention likelihood. The latent classes may also differ with regard to how these propensities change over time (duration dependence), the correlation between acquisition and retention processes (governed by ω_s), the sensitivity to the promotional activity (β_s), and the likelihood that customers ever acquire service (z_s), all of which we flexibly model as segment specific. Thus, the customer-level likelihood given in Equation (12) will depend on the set of segment-specific parameters $\{z_s, \omega_s, \Theta_s, \beta_s, \Phi_s\}$, and the unconditional likelihood is given by

$$L(\mathbf{z}, \boldsymbol{\omega}, \boldsymbol{\Theta}, \boldsymbol{\beta}, \boldsymbol{\Phi}, \mathbf{q} \mid t_h, d_h, c_h^A, c_h^R, T_h, \mathbf{X}(t)) = \sum_{s=1}^{S} q_s P(z_s, \omega_s, \Theta_s, \beta_s, \Phi_s \mid s, t_h, d_h, c_h^A, c_h^R, T_h, \mathbf{X}(t)), \tag{13}$$

where $\mathbf{z}$, $\boldsymbol{\omega}$, $\boldsymbol{\Theta}$, $\boldsymbol{\beta}$, $\boldsymbol{\Phi}$, and $\mathbf{q}$ denote the vectors of segment-specific parameters, and q_s denotes the probability that a customer is in segment s such that $\sum_{s=1}^{S} q_s = 1$. Letting Y_h denote the observed data vector $\{t_h^A, t_h^R, c_h^A, c_h^R, T_h\}$, the overall log-likelihood (LL) is

$$\text{LL}(\mathbf{z}, \boldsymbol{\omega}, \boldsymbol{\Theta}, \boldsymbol{\beta}, \boldsymbol{\Phi}, \mathbf{q} \mid Y_1, \ldots, Y_H) = \sum_{h \in H} \log(L(\mathbf{z}, \boldsymbol{\omega}, \boldsymbol{\Theta}, \boldsymbol{\beta}, \boldsymbol{\Phi}, \mathbf{q} \mid Y_h)). \tag{14}$$

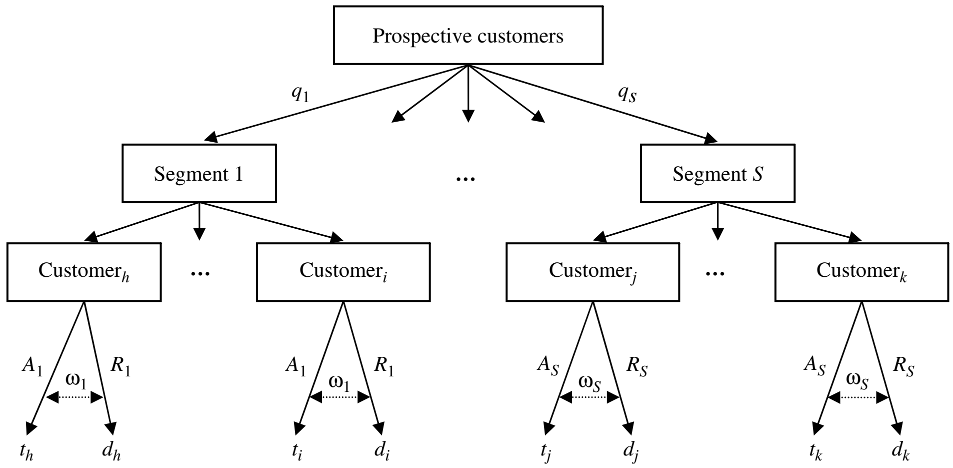
We provide a graphical illustration of the proposed modeling framework in Figure 3, where we denote the acquisition and retention processes for customers in latent class k by A_k and R_k, respectively.

As shown in the first level of Figure 3, customers belong to a particular latent class (or segment). Because all customers in a particular segment follow the same acquisition and retention processes, this introduces the first type of correlation. Given their segment membership, customers' acquisition times and retention durations may be correlated via the Sarmanov family (captured by the value of ω), which provides the second form of correlation.

4.5. Baseline Hazard Specifications

The model presented in Equations (1)–(14) is flexible, allowing for any timing distribution(s) to be specified for the acquisition and retention processes. In our empirical application, we consider three sets of possible baseline hazard specifications for the acquisition and retention processes: the Weibull distribution, log-logistic distribution, and expo-power distribution. Though other baseline hazard specifications can be used, these three are popular choices in extant timing models and each allows for varying forms of duration dependence (e.g., Seetharaman and Chintagunta 2003).

Figure 3 Model Schematic



The Weibull distribution has baseline hazard and survival functions given by

$$h(t \mid \gamma, \alpha) = \gamma \alpha t^{\alpha-1}, \quad S(t \mid \gamma, \alpha) = e^{-\gamma t^{\alpha}}. \tag{15}$$

The Weibull distribution allows for either positive duration dependence (an increasing baseline hazard) when $\alpha > 1$, as well as negative duration dependence (a decreasing baseline hazard) when $\alpha < 1$. The Weibull distribution also nests the exponential distribution ($\alpha = 1$), allowing it to capture the case of no duration dependence.

Whereas the Weibull allows for a flat baseline hazard function or a monotonically increasing or decreasing baseline hazard function the log-logistic distribution, with baseline hazard and survival functions given by

$$h(t \mid \gamma, \alpha) = \frac{\gamma\alpha(\gamma t)^{\alpha-1}}{1+(\gamma t)^{\alpha}}, \quad S(t \mid \gamma, \alpha) = \frac{1}{1+(\gamma t)^{\alpha}}, \tag{16}$$

allows for a baseline hazard that is either monotonically decreasing or has an inverted U-shape.

Last, the expo-power distribution is defined by a baseline hazard and survival function

$$h(t \mid \gamma, \alpha, \theta) = \gamma \alpha t^{\alpha-1} e^{\theta t^{\alpha}}, \quad S(t \mid \gamma, \alpha, \theta) = e^{(\gamma/\theta)(1-e^{\theta t^{\alpha}})}. \tag{17}$$

The expo-power baseline hazard accommodates a variety of shapes, including monotonically increasing or decreasing, inverted U-shaped, and U-shaped.

4.6. An Illustration of Correlated Processes

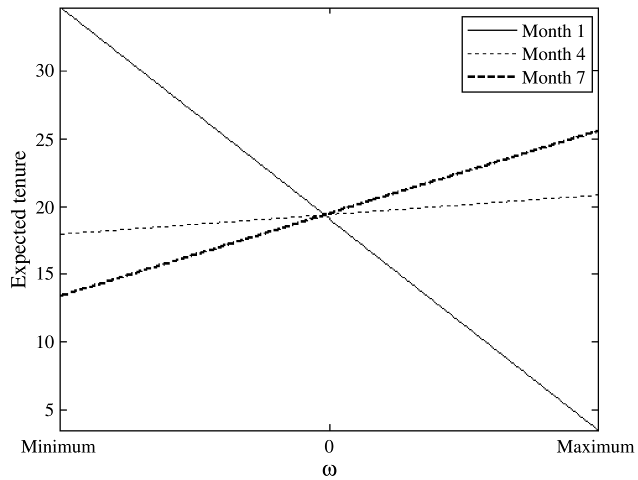
As many readers may be unfamiliar with the Sarmanov family of distributions, before we show our empirical results, we demonstrate via simulation the impact of correlated acquisition and retention processes by computing the expected retention durations conditional on the time of acquisition for different values of ω. Without loss of generality, we use the log-logistic specification for acquisition with baseline hazard $h_A(t_A \mid \gamma = 0.25, \alpha = 1)$, and retention with baseline hazard $h_R(t_R \mid \gamma = 0.5, \alpha = 1)$ and $\beta = -1$. The ET, which is a function of a customer's time of acquisition t_h, is given by

$$\mathrm{ET}(t_h) = \sum_{d=1}^{\infty} \frac{p_2(t_h, d)}{z \cdot f(t_h)}, \tag{18}$$

where the numerator has the same form as $p_2(t_h, d)$ from Equation (10) and the denominator is the marginal probability of acquiring service at time t_h. In Figure 4, we show the ET of service for subscribers who acquire service in months 1, 4, and 7 for the full range of values that ω can assume. For these particular parameter values, the range of ω is from -107 to 111. While this seems to correspond to a narrow range of correlations, the expected retentions (based on time of acquisitions) shown in Figure 3 are dramatically different from each other.

Under these parameters, a customer beginning service after one month may be expected to retain service more than 2.5 times longer than a customer beginning after seven months (at the minimum value of ω), or have an ET more than 80% shorter than that of a customer beginning after seven months (at the maximum value of ω). Note that the direction of the relationship changes as ω changes sign. At $\omega = 0$, the ET of service is independent of acquisition time. Based on the strength and direction of the relationship between acquisition and retention, customers who acquire service at different times may have different ETs and therefore may be of differential value to the service provider.

Figure 4 **Impact of Correlated Outcomes on Expected Tenure for a Simulated Example**



It is also worth noting in Figure 4 that the impact of correlated processes diminishes over time. That is, while the differences in ET are large for early subscribers, this effect is proportionately smaller when comparing later subscribers to each other. This pattern continues to hold for all observed acquisition times and values of ω for multiple sets of parameters that we examined.

In each latent class, therefore, depending on the nature of the correlated relationship as reflected by the value of ω and the type of duration dependence in both the acquisition and retention processes, several behavioral patterns may emerge. For example, a "worst case" scenario for prospect acquisition would include a negative correlation between acquisition and retention and a monotonically decreasing hazard function in the acquisition process. Not only are prospects less likely to acquire the service over time, but those who do acquire service later are expected to discard service faster. In such a situation, firms must recognize that it may not be worthwhile to devote resources to prospects who have not acquired the service after some time.

On the other hand, prospects who have foregone the service for a long period can be valuable under different conditions. An inverse-U shape in the baseline hazard of the acquisition process would indicate that prospects become more likely to acquire service for a period of time, making them candidates for targeting in that period. Coupling this acquisition process with positive correlation between acquisition and retention times, later acquirers will be expected to maintain service for a longer duration than early acquirers. In this case, it may be worthwhile to pursue prospects, at least temporarily.[3] As we will demonstrate in §6, the proposed model therefore can be used to determine the value of prospects and customers.

5. Empirical Analyses

The full model presented in Equations (1)–(14), using the three choices for baseline hazard functions outlined in Equations (15)–(17), was fit using maximum likelihood estimation in MATLAB. For each of the three distributional assumptions, we estimate models with different numbers of latent classes to assess how many classes are needed, based on Bayesian Information Criterion (BIC) (Schwarz 1978). We also estimate a series of models in which we assume that $\omega = 0$, an important nested case that decouples acquisition and retention. In these models, note that we still allow for correlation in the parameters governing the acquisition and retention processes across latent classes, but assume that the processes are conditionally independent. We note that in our analyses we only consider the same distribution for both acquisition and retention (i.e., Weibull for both). It is possible to use distinct distributions to model each process (e.g., Weibull for acquisition and expo-power for retention), but this would require many more models to assess the appropriate number of latent classes and the need for correlated processes. We present the in-sample LL and BIC values for the estimated models in Table 2.

Based on BIC, the log-logistic baseline hazard specification outperforms the expo-power in all cases and beats the Weibull in all cases but one. We next provide a brief comparison of the *best in class* models,

Table 2 **Model Results**

Model	Baseline hazard	Classes	Correlated processes?	No. of Parameters	LL	BIC
1	Log-logistic	1	Yes	7	−7,669	15,398
2	Log-logistic		No	6	−7,676	15,404
3	Log-logistic	2	Yes	15	−7,659	15,449
4	Log-logistic		No	13	−7,659	15,432
5	Log-logistic	3	Yes	23	−7,648	15,496
6	Log-logistic		No	20	−7,649	15,473
7	Weibull	1	Yes	7	−7,715	15,492
8	Weibull		No	6	−7,721	15,495
9	Weibull	2	Yes	15	−7,656	15,444
10	Weibull		No	13	−7,664	15,441
11	Weibull	3	Yes	23	−7,652	15,505
12	Weibull		No	20	−7,655	15,485
13	Expo-power	1	Yes	9	−7,716	15,510
14	Expo-power		No	8	−7,721	15,512
15	Expo-power	2	Yes	19	−7,663	15,491
16	Expo-power		No	17	−7,668	15,485
17	Expo-power	3	Yes	29	−7,649	15,551
18	Expo-power		No	26	−7,653	15,533

[3] The possible behavioral stories that emerge depend not only on the value and magnitude of ω but also on the choice of the baseline hazard function for the acquisition and retention processes.

Figure 5 **Model Performance on Acquisition Process**

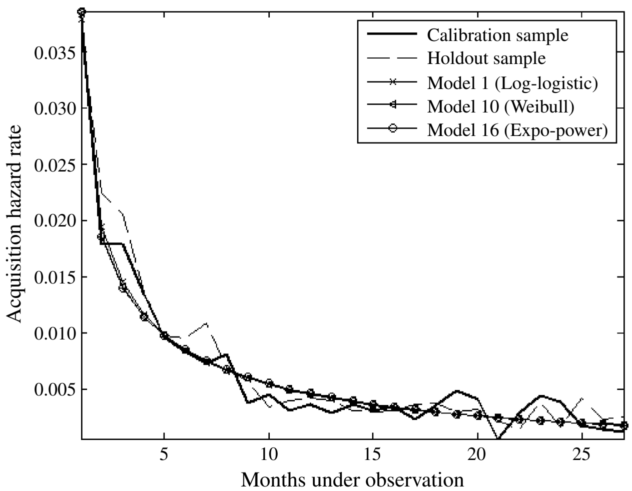
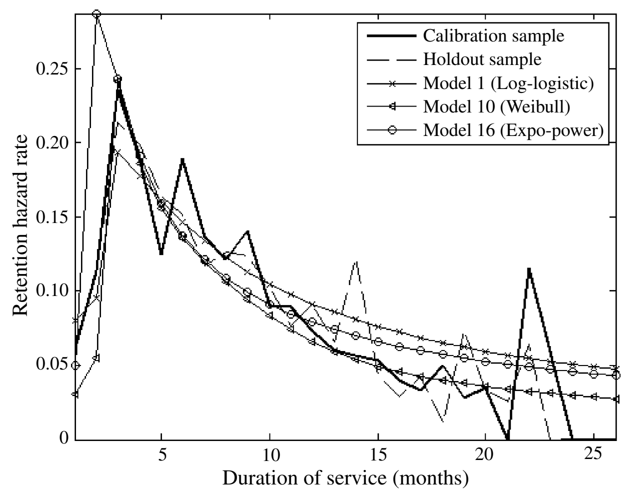


Figure 6 **Model Performance on the Retention Process**



that is, we compare the best performing log-logistic, Weibull, and expo-power models (Models 1, 10, and 16, respectively).

Under the log-logistic specification for acquisition and retention, the best performing model incorporates correlated processes but just a single class of customers (Model 1). This contrasts with the Weibull and expo-power specifications (Models 10 and 16, respectively) that require two latent classes of customers but have conditionally independent processes, highlighting the need to allow for latent classes in the modeling framework. In fact, across the three distributional assumptions, correlated processes (the Sarmanov family) only provide significant improvement (on the basis of BIC) when there is a single class of customers. Thus, while other bivariate timing models have found the need for both correlated outcomes and linked propensities (e.g., Park and Fader 2004), the data used in our empirical application suggests that correlated processes and a single class of customers is adequate to model the relationship between acquisition and retention in our data. Although the BIC indicates that Model 1 seems to be most suitable for this data set, we next extend our best in class analysis to a more detailed analysis of the marginal acquisition and retention processes to demonstrate their overall performance, and then select among them.

To further compare the performance of the three *best in class* models, in Figure 5 we present the empirical hazard rate of the acquisition process for both the calibration and holdout samples, as well as the expected hazard rates.

The expected hazard rates under the *best in class* models all behave similarly. While the observed calibration and holdout hazard rates are fairly jagged, the expected hazard rates under each specification appear to capture the general trend in the observed hazard rate over time. If more time-varying covariates existed, deviations from the smooth trend would be obtainable.[4]

In Figure 6, we show the empirical hazard rates of the retention process for both samples, along with the expected hazard rates from the *best in class* models.

The effect of the promotional activity on the empirical hazard rate is immediately evident, as churn is dampened during the two months following the month of acquisition before dramatically increasing. All three of the *best in class* models reflect this general shape, though the expected hazard rates differ at longer durations. Note, however, that the empirical hazard rate of the calibration sample at these longer durations is based on only a few observations—only 69 (out of 987) customers who acquired service were observed to maintain it for 20 months, and only 4 customers are observed to discard service after 20 months.

Though Figure 5 indicates comparable performance from each of the *best in class* models on the acquisition process, the model predictions from each of the three baseline hazard functions on the retention process differ slightly. The difference between the hazard rates at most durations of service is approximately 0.02. The log-logistic specification has the smallest maximum absolute error with regard to the observed retention hazard. Based on this observation and the BIC values, we choose to present a more detailed view of the results under the log-logistic specification (Model 1) and demonstrate how our framework can be used to value customers.[5]

The parameter estimates governing the log-logistic acquisition process are $\hat{\gamma} = 0.08$ (standard error =

[4] Though one could add a stochastic error term to the acquisition and retention processes, this would eliminate the convenience of a closed-form likelihood function.

[5] Results from the remaining *best in class* models are available in the Technical Appendix online at http://mktsci.journal.informs.org.

0.01), $\hat{\alpha} = 0.71$ (SE $= 0.02$), and $\hat{z} = 0.26$ (SE $= 0.01$) for the split-hazard probability. The parameter estimates governing the log-logistic retention process are $\hat{\gamma} = 0.30$ (SE $= 0.01$), $\hat{\alpha} = 1.31$ (SE $= 0.04$), and $\hat{\beta} = -0.82$ (SE $= 0.07$). As the negative value of β indicates, promotional activity decreases the probability that a current subscriber will discontinue service. Interestingly, the hazard rate, both expected from the model and observed in our two samples, is initially increasing before decreasing. That is, for a short time subscribers are increasingly likely to discard service but later become less likely to discard it.

While Figures 5 and 6 demonstrate the model's performance for each of the marginal processes, it does not shed light on the joint process. We next examine the joint accuracy of the model by comparing its predictions to the data, as originally explored in Table 1. In Table 3, we show the estimated joint distribution of customers who subscribe by the end of the observation period.

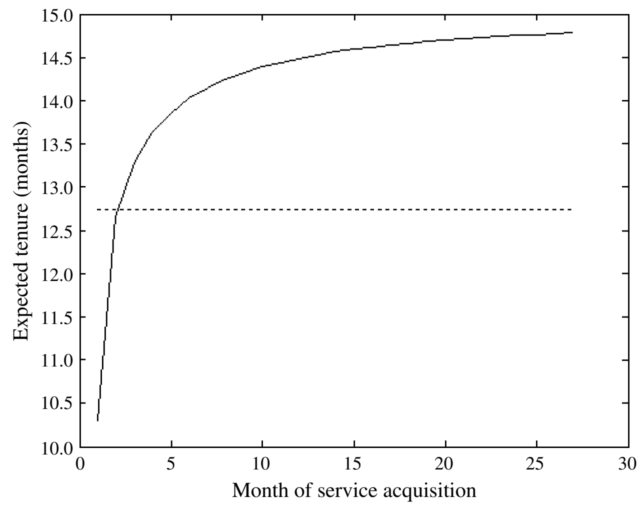
In our calibration sample, 84.1% of the customers do not acquire service during the observation period, which is also predicted accurately by the model. In Table 3, we find that the model closely captures observed acquisition and retention behavior with no more than a 2.6% absolute error in any cell. While only 15.9% of the calibration sample acquires service during the observation period, based on our estimate for z under Model 1, we expect that approximately an additional 10% of customers will eventually acquire service at some point in the future. Thus, even after 27 months of observations (our entire data period), a large number of prospects are still likely to acquire service. This type of insight and forecast requires a formal model that uses the time until acquisition unlike previous methods (e.g., Reinartz et al. 2005, Thomas 2001). The firm can subsequently use such estimates of the service's penetration among prospects to more effectively allocate resources among acquisition and retention activities.

Table 3 **Joint Distribution of Customers Subscribing During the Acquisition Period**

	Retention (%)			
Acquisition	Within 3 months of acquisition	Between 4 and 12 months of acquisition	After 12 months of acquisition	Right censored
Between 1 and 3 months	Observed: 19.3 Predicted: 16.7	O: 17.7 P: 20.1	O: 3.6 P: 4.1	O: 4.5 P: 3.5
Between 4 and 12 months	O: 9.6 P: 11.2	O: 16.5 P: 15.7	O: 1.5 P: 2.5	O: 5.7 P: 5.1
After 12 months	O: 6.3 P: 5.7	O: 4.5 P: 5.4	O: 0 P: 0.1	O: 10.8 P: 9.9

Note. O = observed, P = predicted by Model 1.

Figure 7 **Expected Tenure as a Function of Time of Acquisition**



Our estimate for the covariance parameter ω ($\hat{\omega} = 44.02$, SE $= 9.75$) supports the positive relationship between time of acquisition and duration of service retention, as briefly demonstrated in Table 3.[6] To demonstrate the impact of the correlation between the two processes rather than just report $\hat{\omega}$, we present the ET as a function of the time of acquisition in Figure 7.

The ET for a customer acquiring service after 1 month is 10.3 months, which increases for customers who acquire service in later months. Customers acquiring service after 1 year have an ET of 14.5 months, more than 40% longer than customers acquiring service after 1 month. If we were to ignore the correlation between the acquisition and retention processes, as in Model 2, the ET would be 12.7 months, indicated by the dashed line in Figure 7. Note that under Model 2, neither linked propensities nor correlated outcomes are present and thus the time of acquisition is assumed to be completely independent of the duration of service retention.

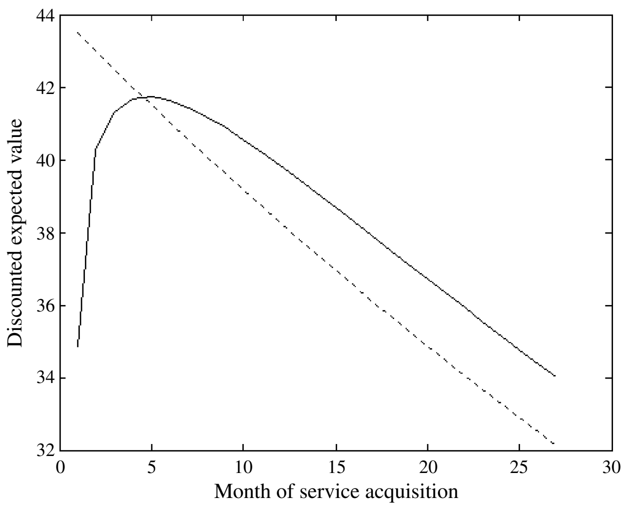
6. Computing Discounted Expected Value for Managerial Decisions

As shown in Figure 7, when we consider the correlation between acquisition and retention, the ET of service subscription is greater for subscribers who acquire service later. Does this imply that later subscribers are always more *valuable* to the service provider? While ET increases for later subscribers, future revenue streams are discounted which may temper their increased tenure.

To assess which customers are the most valuable, we compute the *discounted expected value* (DEV) of

[6] The possible range for ω given the other parameter estimates is -107.9 to 122.6.

Figure 8 Discounted Expected Value



customers based on their time of acquisition. To estimate the DEV, we modify Equation (18) to incorporate a discount factor. The DEV of a customer who acquires service after t months under observation is given by

$$\mathrm{DEV}(t) = \sum_{y=1}^{\infty}\left(\frac{m(t)}{(1+d)^{y+t}}\cdot\frac{p_2(t, y)}{z\cdot f(t)}\right). \tag{19}$$

The summation takes the calculation of ET from Equation (18) for a customer who acquires service in month t, discounts it for each month in the future by a factor of d, and multiplies that by the margin of service $m(t)$.[7]

In Figure 8, we assume a 15% annual discount rate and a margin consistent with the service provider's current pricing to demonstrate how the DEV changes based on the time of acquisition.[8]

Initially, the DEV of subscribers increases with their time of acquisition from a value of $34.85 for those who acquire after just one month to $41.72 for those acquiring in month five, an increase of 20%. Despite the increase in ET with the time of acquisition, the DEV diminishes for prospects who acquire service after five months. Although late-arriving customers have a longer expected tenure, they do not acquire service for several months (or years), thereby diminishing the *present value* of their longer subscriptions.

If we were to erroneously ignore the relationship between time of acquisition and retention, our estimate of DEV (based on Model 2) is reflected by the dashed line in Figure 8, which is monotonically decreasing with the time of service acquisition. While the estimates of DEV with and without the relationship between acquisition and retention eventually mirror each other, early subscribers are overvalued when the relationship is ignored. In particular, a customer who subscribes after one month would have an estimated value of $43.49, which is 25% greater (a very significant amount) than our model-based estimate when the two outcomes are linked, highlighting the need to account for the relationship between acquisition and retention.

This example illustrates the applicability of our proposed framework, as DEV can be incorporated into managerial decisions such as resource allocation and customer targeting (e.g., Rust and Chung 2006). As a measure of the long-run revenue generated by a subscription, it can also be weighed against the costs of acquisition activities and marketing expenditures (e.g., Gupta and Zeithaml 2006). Thus, a provider should be cautious of implementing proposed promotional activities or other initiatives for which the cost exceeds the DEV of the prospects at which they are aimed. Miscalculations of DEV, such as the one illustrated above, could lead the firm to devote more resources than it should to some of its customers. In addition, the resulting joint acquisition and retention curves can be incorporated into the kinds of resource allocation tasks that have been discussed by authors such as Blattberg and Deighton (1996). Our estimate of DEV can also be incorporated into dynamic programming applications. For example, managers can determine, based on the DEV and the acquisition process, when acquisition activities should cease.

7. Conclusions and Implications

We have developed a bivariate correlated split-hazard model of the time until acquisition and duration for which customers retain service. Our model allows us to estimate the fraction of customers who will eventually acquire service and to explore the joint timing process. In the spirit of Park and Fader (2004), we allow for *double correlation* within and across customers for these two processes. While other work has examined the relationship between a binary acquisition process (e.g., a yes/no decision) and subsequent retention, to the best of our knowledge, our work is the first to jointly model the sequential acquisition and retention *spells* in a contractual setting.

Testing for the link between acquisition and retention can have important implications for marketing activities such as customer valuation and resource allocation. If a link exists between the acquisition and retention processes, certain customers may be

[7] This calculation can easily be generalized to accommodate multiple latent classes by incorporating the updated probability of class membership q_s, based on the time of acquisition.

[8] While we assume the costs of providing service to be zero, if such information were available it could easily be incorporated (e.g., Gupta and Zeithaml 2006).

more valuable to the firm than others, as demonstrated by Figure 8. In our application, we find a positive relationship between the time of acquisition and subsequent duration of service for those who acquire service. While those who acquire service later tend to keep service for longer durations, this benefit must be weighed against the time (and cost) it takes to acquire these prospects. If pricing information and the cost of acquisition efforts were available, our measure of DEV could be incorporated into a break-even analysis to determine if and when the firm should cease acquisition efforts.

There are certain limitations to our proposed methodology that must be acknowledged. First, our data covers only one service for one company; as such, it is premature to generalize our findings to other applications. In addition to considering a wider range of services, future research should also consider the interplay of competing services and/or firms in the acquisition and retention of service with a particular provider. Popular notions such as *share of wallet* require extensions to the multiservice and multifirm setting.

We find support for a single class of customers who exhibit correlated acquisition and retention behavior. Alternative baseline hazard specifications, however, yield results in which multiple classes of customers exhibit conditionally independent acquisition and retention behavior within their classes. Data that is not as coarse may enable for greater differentiation between these competing stories, as authors using more granular data (e.g., Park and Fader 2004) have found support for both types of correlation. Based on Figures 5 and 6, which reveal that the alternative model specifications perform similarly (especially with regard to the acquisition process), a latent class approach and a model of correlated processes may be able to approximate each other. As such, additional research is warranted in assessing the limitations imposed by discrete-time data on modeling correlated behavior, as well as more clearly assessing the type of correlation that exists in the data.

Future work should also introduce other marketing variables and customer characteristics to which we did not have access. This would allow the modeling framework to be used to assess the implications of different policies, as well as the value of targeting customers with specific actions (e.g., Palmatier et al. 2006). If additional information about customers were available, such as customers' expected future usage of the service (e.g., Lemon et al. 2002), this could be linked to both the acquisition and the retention processes. It would also be useful to determine if the mode of acquisition has any impact on the acquisition-retention relationship. In addition, models must be developed to determine what aspects of behavior are impacted by marketing activities. While some types of intervention may affect acquisition or retention (or both), they may also affect the relationship between the two processes (e.g., allow ω to be a function of $\mathbf{X}(t)$). It might also make sense to investigate whether there is a dynamic aspect to this relationship: Does the acquisition-retention association evolve over time (ω as a function of t) as a service (and its customer base) matures? The future pool of customer prospects might have different acquisition-retention tendencies than the current group. These tendencies may also be influenced by the behavior of current and lapsed customers via word-of-mouth.

It is too early to speculate about the existence and nature of such dynamic relationships. We do hope that other researchers will continue down this path to better understand the interplay among the underlying processes that new customers follow as they come to a new service provider and subsequently depart. It is not clear what kinds of substantive observations or "empirical generalizations" will arise, but it is important for researchers to use the right tools to uncover and characterize them in future investigations.

Acknowledgments
The authors thank the anonymous firm that graciously provided the data used here.

Appendix. Derivation of Mixing Functions

A general form of the mixing function for the Sarmanov family of bivariate distributions is given in Equation (5). Rewriting it to account for the discrete-time data, we replace the integral with a summation over all periods of time

$$\phi(x) = f(x) - \sum_t f^2(t). \tag{20}$$

For the acquisition and retention processes, the discrete-time probability mass function is given by

$$f(t) = S(t-1) - S(t) \quad \text{for } t = 1, 2, 3, \ldots, \tag{21}$$

where $S(t)$ is the survival function at time t. Substituting Equation (21) into Equation (20), we find that

$$\phi(x) = (S(x-1) - S(x)) - \sum_{t=1}^{\infty} (S(t-1) - S(t))^2. \tag{22}$$

Expanding Equation (22) yields

$$\begin{aligned}\phi(x) &= (S(x-1)-S(x)) \\ &\quad - \left(\sum_{t=1}^{\infty}(S(t-1))^2 - 2\sum_{t=1}^{\infty}(S(t)S(t-1)) + \sum_{t=1}^{\infty}(S(t))^2\right),\end{aligned} \tag{23}$$

which can be rewritten as

$$\begin{aligned}\phi(x) &= (S(x-1) - S(x)) \\ &\quad - \left(\sum_{t=0}^{\infty}(S(t))^2 - 2\sum_{t=1}^{\infty}(S(t)S(t-1)) + \sum_{t=1}^{\infty}(S(t))^2\right),\end{aligned} \tag{24}$$

and then as

$$\phi(x) = (S(x-1) - S(x)) - \left(1+\sum_{t=1}^{\infty}(S(t))^2 - 2\sum_{t=1}^{\infty}(S(t)S(t-1)) + \sum_{t=1}^{\infty}(S(t))^2\right). \quad (25)$$

Further simplifying Equation (25), we derive the final form of the mixing function

$$\phi(x) = (S(x-1) - S(x)) + 2\left(\sum_{t=1}^{\infty} S(t)((S(t-1) - S(t)))\right) - 1. \quad (26)$$

References

Berger, P. D., N. Nasr-Bechwati. 2001. The allocation of promotion budget to maximize customer equity. *Omega* **29**(1) 49–61.

Bitran, G. R., S. V. Mondschein. 1996. Mailing decisions in the catalog sales industry. *Management Sci.* **42**(9) 1364–1381.

Blattberg, R. C., J. Deighton. 1996. Manage marketing by the customer equity test. *Harvard Bus. Rev.* **74**(4) 136–144.

Chintagunta, P. K., S. Haldar. 1998. Investigating purchase timing behavior in two related product categories. *J. Marketing Res.* **35**(1) 43–53.

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

Danaher, P. J., B. G. S. Hardie. 2005. Bacon with your eggs? Applications of a new bivariate beta-binomial distribution. *Amer. Statistician* **59**(4) 282–286.

Farlie, D. J. G. 1960. The performance of some correlation coefficients for a general bivariate distribution. *Biometrika* **47**(December) 307–323.

Follman, D. A., M. S. Goldberg. 1988. Distinguishing heterogeneity from decreasing hazard rates. *Technometrics* **30**(4) 389–396.

Gupta, S., V. Zeithaml. 2006. Customer metrics and their impact on financial performance. *Marketing Sci.* **25**(6) 718–739.

Gupta, S., D. R. Lehmann, J. A. Stuart. 2004. Valuing customers. *J. Marketing Res.* **41**(1) 7–18.

Hansotia, B. J., P. Wang. 1997. Analytical challenges in customer acquisition. *J. Direct Marketing* **11**(2) 7–19.

Jain, D., S. Singh. 2002. Customer lifetime value research in marketing: A review and future directions. *J. Interactive Marketing* **16**(2) 34–46.

Johnson, N. L., S. Kotz. 1975. On some generalized Farlie-Gumbel-Morgenstern distributions. *Comm. Statist.* **4**(5) 415–427.

Johnson, N. L., S. Kotz. 1977. On some generalized Farlie-Gumbel-Morgenstern distributions—II: Regression, correlation, and further generalizations. *Comm. Statist.: Theory Methods* **A6**(6) 485–496.

Kamakura, W., G. Russell. 1989. A probabilisitic choice model for market segmentation and elasticity structure. *J. Marketing Res.* **26**(November) 379–390.

Kamakura, W. A., B. S. Kossar, M. Wedel. 2004. Identifying innovators for the cross-selling of new products. *Management Sci.* **50**(8) 1120–1133.

Kotz, S., N. Balakrishnan, N. L. Johnson. 2000. *Continuous Multivariate Distributions, Volume 1: Models and Applications, Second Edition*. John Wiley and Sons, New York.

Lee, M.-L. T. 1996. Properties and applications of the Sarmanov family of bivariate distributions. *Comm. Statist.: Theory Methods* **25**(6) 1207–1222.

Lemon, K. N., T. B. White, R. S. Winer. 2002. Dynamic customer relationship management: Incorporating future considerations into the service retention decision. *J. Marketing* **66**(January) 1–14.

Milne, G. R., A. Rohm. 2000. Consumer privacy and name removal across direct marketing channels: Exploring opt-in and opt-out alternatives. *J. Public Policy Marketing* **19**(2) 238–249.

Morrison, D. G., D. C. Schmittlein. 1980. Jobs, strikes, and wars: Probability models for duration. *Organ. Behav. Human Performance* **25**(2) 224–251.

Palmatier, R. W., S. Gopalakrishna, M. B. Houston. 2006. Returns on business-to-business relationship marketing investments: Strategies for leveraging profits. *Marketing Sci.* **25**(5) 477–493.

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

Reinartz, W., J. S. Thomas, V. Kumar. 2005. Balancing acquisition and retention resources to maximize customer profitability. *J. Marketing* **69**(January) 63–79.

Rust, R. T., T. S. Chung. 2006. Marketing models of service and relationship. *Marketing Sci.* **25**(6) 560–580.

Schwarz, G. 1978. Estimating the dimension of a model. *Ann. Statist.* **7**(2) 461–464.

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

Sinha, R. K., M. Chandrashekaran. 1992. A split hazard model for analyzing the diffusion of innovations. *J. Marketing Res.* **29**(1) 116–127.

Thomas, J. S. 2001. A methodology for linking customer acquisition and customer retention. *J. Marketing Res.* **38**(2) 262–268.