

Modeling Customer Lifetimes with Multiple Causes of Churn

Michael Braun
MIT Sloan School of Management
Massachusetts Institute of Technology
Cambridge, MA 02139
braunm@mit.edu

David A. Schweidel
Wisconsin School of Business
University of Wisconsin - Madison
Madison, WI 53706
dschweidel@bus.wisc.edu

September 3, 2010

Abstract

Customer retention and customer churn are key metrics of interest to marketers, but little attention has been placed on linking the different reasons for which customers churn to their value to a contractual service provider. In this article, we put forth a hierarchical competing risk model to jointly model when customers choose to terminate their service and why. Some of these reasons for churn can be influenced by the firm (e.g., service problems or price-value tradeoffs), but others are uncontrollable (e.g., customer relocation and death). Using data from a provider of land-based telecommunication services, we examine how the relative likelihood to end service due to different reasons shifts during the course of the customer-firm relationship. We then show how the effect of a firm's efforts to reduce customer churn for controllable reasons are mitigated by presence of uncontrollable ones. The result is a measure of the incremental customer value that a firm can expect to accrue by delaying churn for different reasons. This "upper bound" on the return of retention marketing is always less than what one would estimate from a model with a single cause of churn and depends on a customer's tenure to date with the firm. We discuss how our framework can be employed to tailor the firm's retention strategy to individual customers, both in terms of which customers to target and when to focus efforts on reducing which causes of churn.

1 Introduction

Customer retention continues to be a topic of importance to marketing researchers and managers. Much of the recent research on customer retention has linked retention rates and churn probabilities to forecasts of customer lifetime value (Fader and Hardie 2010), managing customer equity (Rust et al. 2004), balancing the allocation of resources among retention and other marketing efforts (Reinartz and Kumar 2003; Reinartz et al. 2005), and financial reporting and management (Gupta et al. 2004). With an interest in managing retention, influencing it with marketing actions, and understanding the effect of retention on the value of the customer base, the focus has primarily been on *when* customers terminate a relationship. That is, the approach has been to model the time until churn, or the duration of the customer relationship. In this sense, the reason for which customers decide to discard service service is irrelevant. All that matters is that they have churned after a particular duration, marking the end of that customer's revenue stream.

A separate stream of research has focused on *why* customers choose to cancel service or switch to other service providers. There may be a number of reasons for why customers ultimately churn. Some customers may churn due to service failures or the actions of competitors. Others may cancel service for reasons beyond the control of the firm, such as the customer relocating to another city, a change in personal circumstances unrelated to the product, or even death. Keaveney (1995) conducted an investigation of how certain critical incidents caused customers to switch from one service provider to another. She surveyed more than 500 service customers, producing a list of more than 800 critical behaviors on the part of the firm. These behaviors were classified into eight categories of critical incidents, including pricing, core service failures, service encounter failures, employee responses to service failures, and competition. Each of these is a potential cause of churn that the firm can potentially influence, at least to some extent. The most cited reason for customers' decisions to switch service providers was "core service failures," which included mistakes, billing errors, and service problems. "Service failure encounters," "pricing," and "employee response to service failures," were second, third, and fourth most listed reasons. Though most of the reasons for which customers churn can be influenced by the firm, as Keaveney (1995) points out, there are some reasons for churn which they cannot control. Bogomolova and Romaniuk (2009) surveyed 765 business-owners who canceled their electronic funds transfer services with an Australian bank, and found that 57 percent of the cancelations were for reasons outside the control of the bank. Sharp (2010) includes several examples of products and services for which some customers defect for reasons that the firm could never reasonably prevent.

In addition to looking at the events that trigger churn, this stream of work has also examined how customers' attitudes relate to their decisions to retain service. Bolton (1998) considers the influence of

customer satisfaction on the duration for which customers choose to maintain service. Verhoef (2003) examines the role of affective commitment and marketing actions on retention. As expected, higher affective commitment is found to be positively related to customer retention. He also reports that membership in loyalty programs has a positive effect on customer retention, as does Bolton et al. (2000). Gustafsson et al. (2005) study the effect of customer satisfaction, commitment, and changes in the basis of the relationship on service retention. Though previous research has identified the drivers of customer churn, including the impact of satisfaction, social bonding and reactional triggers (Oliver 1999; Gustafsson et al. 2005), the reasons for which customers terminate a relationship with a firm have yet to be incorporated into quantitative models of customer lifetime or residual customer valuation.

This raises a key question: although there is evidence that the reason for churn may vary from customer to customer, does it even matter? From the perspective of a shareholder of the firm who is evaluating the financial health of the organization, the answer may be “no.” The firm’s aggregate churn (or retention) rate may be sufficient for assessing the value of the customer base in terms of the stream of discounted revenue that the firm can expect to accrue from the existing customer base (Gupta et al. 2004; Fader and Hardie 2010).

Consider, however, the role of a manager who has the operational responsibility for maintaining those revenue streams. Managers require an understanding of not just when customers churn, but also why they churn. With such information, managers can respond accordingly, by pulling the appropriate marketing and operational “levers” that might forestall when a customer churns for a particular reason. For example, if customers cancel primarily due to service failures, the firm might prioritize investments that focus on improving the quality of service (Rust et al. 1995). On the other hand, if customers are switching to service providers who are seen to offer a more appealing product for a better price, the firm might instead engage in marketing activities that emphasize their own perceived value proposition.

Although these reasons for churn can, at least to some extent, be influenced by the firm’s actions, what about causes of churn beyond its control, such as relocation or death? No amount of marketing investment is likely to have any effect in delaying churn occurring due to such circumstances. Moreover, if these uncontrollable causes of churn tend to be more prevalent among customers than those causes that the firm can influence (i.e., “controllable” churn), then the impact of investing in reducing controllable churn may be limited substantially. More generally, the presence and prevalence of each cause of churn, which may vary across customers and over the course of a customer’s relationship, will limit the effectiveness of efforts to delay churn that occurs due to all other causes. Incorporating the reasons for churn into forecasting and analytical infrastructure can not only provide useful insights into which customers are likely to churn, at what times, and for what reasons, but can also help the firm determine how much return it can expect to

recoup from its customer retention efforts.

In this article, we draw both on research that has examined the underlying factors that contribute to customer churn and on research on constructing probability models of customer retention and lifetime, incorporating customers' stated reasons for terminating service into the model structure. Extant duration models of customer retention in a contractual relationship assume that the time at which a customer decides to churn is governed by a single stochastic process and that customers vary in their underlying propensities to churn at any particular time¹. Rather than assuming that customer churn is a single process, we consider the different reasons for which customers may churn and characterize the propensities for them to cancel service for each of these reasons according to separate processes. When a customer eventually churns, he is canceling service for exactly one of these reasons, consequently rendering the other reasons for churn irrelevant. As an analogy, there are many events or ailments that may cause a patient's death, but his death is caused by whichever reason gets him first. In short, we employ a competing risk model at the individual level, where different reasons for churn (henceforth referred to as *risks*) race to decide which risk will trigger the eventual churning act.

Our *hierarchical* competing risk model allows for both observed and unobserved heterogeneity across the customer base in the expected churn propensities for each specific risk. Competing risks models are not new to the marketing literature, as we discuss in Section 2. However, the competing risk structure lets us simultaneously model the relationship between time of churn and reason for churn, yielding novel insights for researchers and practitioners interested in customer retention. Using data from a provider of land-based telecommunications services, we estimate and evaluate a particular competing risk model specification. This dataset includes the length of time that service was maintained for all customers, including those who churned during the observation period or survived throughout it. It also includes a customer-stated reason for churn for most customers who canceled service, and some geo-demographic information for both churning and surviving customers. Such information is often available to contractual service providers, as customers must contact the firm in some way to terminate the relationship. To account for those customers who choose not to provide a reason for why they are terminating service, we treat this as a missing data problem and employ standard methods to address it (see Section 3.1).

Linking these two streams of research generates a number of interesting findings. As detailed in Section 3, we find that for some modeling and prediction tasks, such as forecasting time until churn regardless of the reason, a single risk duration model of customer lifetime (even without any demographic information) does about as well as our competing risk model. Unobserved variations among risk-specific churn propen-

¹By "contractual," we mean that the time and event of churn is directly observed, in the same sense as Fader and Hardie (2009) and Bolton (1998).

sities are simply absorbed into the population-level mixing distribution, and such heterogeneous models may be flexible enough to predict the time of churn. It is when we want to understand the proportion of customers who churn due to a particular risk, be it one that the firm may influence or not, and how those proportions evolve as the customer's tenure with the firm progresses, that a competing risk model starts to shine. As the likelihood of churn due to each individual reason may change as a customer "ages," the relative propensities of churn due to different risks vary. The competing risk framework is flexible enough to capture this dynamic, as well as the extent of churn regardless of risk, even for a subset of customers whose data was not involved in estimating the population-level parameters. Thus, we establish that a hierarchical competing risk model is a reasonable and parsimonious way to represent joint time-to-churn and cause-of-churn behavior.

Having shown that the proposed competing risk model captures the trends in both of these aspects of the observed data, in Section 4 we discuss the managerial insights about customer retention and expected lifetimes that are afforded by our analysis. Our key contribution is in how we leverage the competing risk framework to generate a more nuanced view of the role that customer heterogeneity plays in understanding the duration of contractual customer relationships than existing models of customer retention can provide. We find that there is considerable variation across geo-demographic clusters in their risk-specific propensities to discard service, suggesting that the incremental benefit of delaying churn due to a particular risk may vary across these clusters. To demonstrate this, we illustrate some patterns in the returns that firms can expect to accrue by investing in slowing customer churn due to specific risks for different geo-demographic clusters. Depending on the relative propensities of different risks after a particular duration in a customer's relationship, some of which are beyond the firm's control, we show that slowing churn due to reasons that the firm can control may in fact do very little to increase an existing customer's remaining value. This is due to a dampening effect imposed by the prevalence of the causes of churn that are beyond a firm's control. The strength of this dampening effect, however, may diminish over time as the likelihood of churn due to different causes changes.

Though intuitive in nature, this dampening effect has not been discussed previously in the literature that has linked customer retention to lifetime and value (Gupta et al. 2004; Fader and Hardie 2010). Consequently, after customers have maintained service for an extended period of time and those reasons for churn that the firm can influence become more likely, the firm may see larger returns on its investments. If a firm were to examine changes in retention at the aggregate level, they may erroneously conclude that their actions are having little effect on customer retention. On the contrary, their actions aimed at curbing customer churn may be quite effective, but only for a subset of customers who are likely to discard service for a reason that the firm can actually influence. As the impact of the firm's actions depend on a customer's

tenure to date, this suggests that retention efforts should be dynamically tailored to individual customers, rather than applied across the board (to all customers) or continually throughout customers' relationships. Even if the overall retention rate, either of an individual customer (Fader and Hardie 2010) or the entire customer base (Gupta et al. 2004; Blattberg et al. 2001), appears constant, digging deeper into how the reasons for churn evolve over a customer's tenure may reveal that returns on the firm's retention efforts vary with customer tenure.

This has clear implications for resource allocations and is an important caveat given the relationship between increased customer retention and customer value that has been documented in the marketing literature. In essence, there may be only so much that a firm can do to increase the likelihood of service retention for an individual at a given time. Just as a brand may not be able to drive loyalty to 100% because of heterogeneity in individuals' brand preferences (Ehrenberg et al. 2004; Sharp 2010), a service provider cannot completely stop customer churn because of the different reasons for which customers will eventually discard service (Blattberg et al. 2001). Being able to estimate how risk-specific churn probabilities change over the course of a customer's relationship with the firm can aid managers deciding both *when* to allocate retention and service improvement resources, as well as to *which* of the firm's existing customers (Venkatesan and Kumar 2004). That is, a firm's customer management strategy does not just vary across individuals, but also over time based on expectations of a customer's *remaining* value. Armed with knowledge of how marketing and service investments can reduce churn due to different causes, firms can then link these investments to changes in customer lifetime to justify those expenditures and develop a dynamic strategy that lets them make best use of limited resources (Hogan et al. 2002; Rust et al. 2004; Reinartz et al. 2005).

2 A Competing Risk Model of Customer Retention

Duration models have been used commonly to relate customers' propensities to retain service to their lifetime value. Berger and Nasr (1998) illustrate how to calculate customer lifetime value based on customers' retention probabilities. Schweidel et al. (2008b) model the retention of contractual services from a telecommunications provider using a Weibull distribution, which parsimoniously and flexibly allow for duration dependence in customers' propensities to churn. Using a proportional hazard model, the authors incorporate both calendar effects and cohort effects. They note that, while cohort effects can improve the fit of the model during calibration, such effects often result in poorer out-of-sample fit.

Fader and Hardie (2010) employ the shifted beta-geometric distribution to model customer retention. The authors demonstrate the importance of including customer heterogeneity in modeling customer value,

finding that the addition of unobserved heterogeneity yields retention elasticities that are higher than those computed by using an aggregate retention rate, as in (Gupta et al. 2004). That is, in ignoring heterogeneity across customers, firms may underestimate the impact of increasing customer retention. These retention elasticities vary as a function of the underlying distribution of retention rates across customers.

While distinct models have been developed for contractual and noncontractual relationships (Fader and Hardie 2009), Borle et al. (2008) develop a customer lifetime value model for a context that contains elements of both contractual and noncontractual settings. Using data from a membership-based direct marketing company, the authors employ a duration model for the time until membership is cancelled. The retention model is then estimated jointly with models of inter-purchase time and purchase amount (the noncontractual elements), using a hierarchical Bayesian model to allow for correlation among the parameters of the three model components.

Though previous research has demonstrated the importance of allowing for heterogeneity across customers, these models treat all churn as the same, and consider only the time at which it occurs. But each reason for churn has its own likelihood of occurring, which may vary not only across the customer base, but also over the course of the customer's relationship with the service provider. To incorporate this into our analysis and subsequent illustrations of the applicability of our modeling framework, we jointly model the duration after which churn occurs and the reason for which a customer churns using a competing risk model.

Competing risk models are often employed when the observed data includes both the time of an event and the cause of that event. Hoel (1972) first modeled competing risks using latent lifetimes, and Prentice et al. (1978) introduced the hazard rate approach that we discuss later in this section. The competing risk framework can thus be considered a joint model for data consisting of a duration outcome (when does an event occur) and a multinomial choice outcome (which cause triggered the event). We refer to these causes as "risks" because they represent the forces that place the individual at risk of the event (we use the terms "cause," "risk," and "reason for churn" interchangeably). Competing risk models are common in medical fields such as epidemiology (Putter et al. 2007), where there are multiple potential causes of death but only one may be observed. Competing risk models have also been used in analyzing unemployment (Han and Hausman 1990).

The competing risk framework has been applied to a number of contexts within the marketing literature. Vilcassim and Jain (1991) and Chintagunta (1998) employ a competing risk setup to jointly model inter-purchase time and brand switching behavior using a continuous-time semi-Markov model. The likelihood of the model is comprised of the likelihood of a particular brand being chosen after an observed duration, as well as the likelihood that all other brands are not chosen before that time. Srinivasan et al. (2008) use

a competing risk model to examine the survival of high tech firms. In their analysis, they consider the different ways in which firms may exit the industry, either by dissolution or by being acquired. Moe and Trusov (forthcoming) model the process by which product reviews are posted on a website by assuming that ratings of different levels each arrive according to their own process, which may be influenced by the previously posted ratings. Though the necessary data is often collected by contractual service providers, to the best of our knowledge the competing risk framework has not been employed in examining the nuances of customer lifetimes.

When there is only one risk to consider, the competing risk model reduces to the standard duration models as a special case, such as those described previously. For a stochastic timing model with a single risk, we define t_i as the elapsed time from when person i (out of a population of size N) signs up for a service to when he either cancels the service, or the end of the observation period of duration T_i , whichever occurs first. Also, define d_i as a non-censoring indicator. If i cancels the service before T_i , then t_i represents i 's lifetime as a customer, and $d_i = 1$. But if i does not cancel before T_i , then i 's lifetime is right-censored and $d_i = 0$. Thus, the observed data for person i is the vector $[t_i, d_i]$. Denote as $F(t_i | \theta_i)$ as the probability that person i churns before time t_i , even if the data weren't censored at T_i , conditioning on θ_i , which is a parameter vector of length r . The choice of $F(t_i | \theta_i)$ leads to functions of θ_i that represent our beliefs about values of interest related to the individual's lifetime. For example, if $F(t_i | \theta_i) = 1 - e^{-\theta_i t_i}$ (an exponential distribution), then $1/\theta_i$ is that person's expected time to churn.

Competing risk models generalize such models by allowing for multiple causes of churn. The intuition is that churn could potentially be triggered by one of several different risks, with the time of churn due to a particular risk being governed by a *risk-specific* random process. However, we can only observe churn due to a single risk—whichever risk first causes it. As such, once customer i churns for risk j , all other risks are rendered irrelevant because i has already churned. The observed data vector includes not only t_i and d_i , but also j_i , an index from 1 to J that identifies which of the J possible risks is the one that i reports as being the reason for churn (j_i is irrelevant for censored observations). In essence, we model observed customer lifetimes as being determined by a race among J churn processes, where the observed cause of churn is risk j_i , and where $d_i = 1$ because the race has finished before time T_i . If i does not churn before the end of the observation period (the data is right-censored), then $d_i = 0$. Under such circumstances, j_i is undetermined, since i has survived all risks for a duration T_i .

The most important notational difference between the single-risk and competing-risk models is that there are now J risk-specific timing distributions, $F_j(t_i | \theta_{ij})$, each controlled by its own parameter θ_{ij} . While we assume that the parameters for the churn process of one risk do not directly influence the churn processes for the other risks, we will allow for correlation among the risk-specific parameters. This correlation

allows for the possibility that some individuals are simply more prone to churn, regardless of the risk, compared to other customers. The resulting parameter vector θ_i is now J times as long as under a single-risk model; it includes the θ_{ij} for each specific risk. Also, note that if a customer does churn because of risk j at time t_i , it means that he has also survived all other risks up to time t_i . So now, $F(t_i|\theta_i)$ is the probability that i churns before time t_i from *any* risk. For our empirical application, we assume that $F_j(t_i|\theta_{ij})$ has the same functional form for all i and j , allowing for differences across the risks through variation in θ_{ij} . However, the competing risk framework could potentially allow for variation in the functional form of $F_j(\cdot)$ across risks.

From these building blocks, the data likelihood for the basic competing risk model is derived easily using hazard rates. In continuous time, a hazard rate $H(t|\cdot)$ is defined as the instantaneous rate of churn at time t , given that it has not happened yet. More formally, $H(t) = f(t)/S(t)$. It is also well-known that there is a one-to-one relationship between a distribution's hazard rate function and its distribution function (and thus, its density and survival functions):

$$S(t|\cdot) = \exp \left[- \int_0^t H(u|\cdot) du \right] \quad (1)$$

In many contractual service contexts such as that used for our empirical application, we must account for interval censoring. This may occur because service is provided through the period for which payment was last submitted or because of the way in which the time of churn is recorded (e.g., monthly). In the case of a single risk, the probability of churning in the t^{th} period is the probability of surviving up to the $(t-1)^{\text{th}}$ period, but not surviving past the t^{th} period. The probability of not churning at all during the observation period is equivalent to surviving past time T . Consequently, the likelihood contribution of a single individual, under a single-risk model, is given simply as:

$$L(t_i|\theta_i) = [S(t_i - 1|\theta_i) - S(t_i|\theta_i)]^{d_i} S(T_i|\theta_i)^{1-d_i} \quad (2)$$

To generalize this to the setting in which there are multiple risks, the data likelihood is the probability that the customer survives *all risks* up to time $t-1$, and then churns because of risk j during the t^{th} period. A customer who survives has survived all risks through time T . Thus, the data likelihood in a competing risk model is

$$L(t_i, j_i|\theta_i) = [(S_j(t_i - 1|\theta_{ij}) - S_j(t_i|\theta_{ij})) S(t_i - 1|\theta_i)]^{d_i} S(T_i|\theta_i)^{1-d_i} \quad (3)$$

Using the risk-specific hazard rates, we can derive the aggregate hazard rate (the instantaneous rate of

churn from *any* risk), which allows us to calculate the aggregate survival probabilities $S(t|\theta_i)$ and the data likelihood in Equation (3). When there is more than one possible risk, the aggregate hazard rate $H(t|\theta_i)$ at time t is simply the sum of the risk-specific hazard rates $H_j(t|\theta_{ij})$ across different risks (Prentice et al. 1978).

$$H(t|\theta_i) = \sum_{j=1}^J H_j(t|\theta_{ij}) \quad (4)$$

A generally-applicable way to derive the aggregate survival probability is to plug (4) into (1). Alternatively, when the hazard rate depends only on the elapsed time of the processes, we can treat the aggregate survival probability as the joint probability of risk-specific survival probabilities.

$$S(t|\theta_i) = \exp \left[- \sum_{j=1}^J \log S_j(t|\theta_{ij}) \right] \quad (5)$$

The data likelihood depends only on the specification of the risk-specific timing models and the person- and risk-specific parameters θ_{ij} for those models.

Without any modification or ad-hoc model specification, the competing risk framework facilitates examining how the relative propensity to churn for the various risks may change over time. If one were to assume a constant hazard rate for each risk, θ_{ij} , then the probability of churning due to a particular risk is given by the ratio $\theta_{ij} / \sum_{k=1}^J \theta_{ik}$. More generally, the relative risk of churning due to risk j at time t can be expressed as the ratio of the hazard rates: $H_j(t|\theta_{ij}) / \sum_{k=1}^J H_k(t|\theta_{ik})$ (Beyersmann et al. 2009). When the risk-specific hazards change over time (i.e., in the presence of duration dependence), the probability with which a customer churns due to a particular cause will also evolve. Such changes in the relative risks were illustrated in a brand switching context by Vilcassim and Jain (1991) and Chintagunta (1998). In the case of customer churn, this may suggest that firms can benefit by employing different marketing activities for different customers, depending on a customer's elapsed tenure with the firm. If the reason for which a veteran customer is likely to discard service compared to a recently acquired customer, the firm may shift its retention strategy with customers' service durations, or decide that their resources are best spent on other customers.

3 Empirical Example: A Telecommunications Service Provider

3.1 The Data

To illustrate the use of competing risk models as a tool for modeling customer churn and lifetimes, we use data that was provided to us, under confidentiality agreements, from a US-based telecommunications provider. This firm serves customers with a wired, subscription-based service, for which customers are billed monthly from when the customer signs up until when he cancels (churns). The population of potential customers in our dataset consists of all private residences in a set of contiguous suburban cities in a top-20 metropolitan area, who signed up for service from January, 2007 through March, 2008. Our observation period for these homes continues through June, 2008. The service is geographically based (service is brought into a home directly), with one other direct competitor in the area. For each household, we know the month in which service was initiated, and, if the customer churned during the observation, when it was canceled.

The company also provided us with geographically based demographic cluster information for each household. Each household is described as being in one of 66 clusters, which we characterize with seven factors (with levels): urbanicity (urban, suburban or rural), income (low, middle or high), age (low, middle or high), children (yes or no), homeowner (yes or no), employment level (retired, blue collar, professional/management, or white collar/service) and education level (some high school, high school grad, some college, college grad, or grad plus). After removing baseline levels, we were left with 15 geo-demographic variables for each household. We should note that we do not know for sure whether any particular household is described perfectly by its demographic vector; types are inferred by geographic information like census tract. Consequently, inferences about the marginal effect of any geo-demographic variable should keep this in mind. As such data are commonly available and employed by service providers, we employ it in our analysis to highlight the managerial applicability of this research.

For about 77 percent of those customers who did cancel service during the observation period, we also have a “reason for churn.” In aggregate, the dataset includes 30 distinct codes for when explaining why service was canceled. Some of these codes are self-reported (e.g., “poor service”²), and others are determined more objectively (e.g., “nonpayment of bill”). For simplicity, we classified these reasons into three “risks:” Value, Personal, and Non-pay/Abuse (NPA). The Value risk includes anything related to the relative value proposition of the company, such as defecting to a competitor, response to poor service, or the price of the service. However indirectly, we consider all of these reasons to be under the control of the company, and are described as “controllable churn” in the spirit of Keaveney (1995). The Personal

²For better or worse, we always assume that customers are telling the truth.

risk includes all reasons that are outside the control of the company, such as customer death, illness, or moving out of a service area. Since it is unreasonable to expect that any activity on the part of the firm could prevent these reasons for churn, we consider them as “uncontrollable churn.” The NPA risk includes causes related to nonpayment, abuse or theft of service, or other cases in which the firm made the churn decision for the customer. It is not clear whether any of the NPA causes are controllable or not, and of course, some may have actually been voluntary (e.g., a customer dislikes the service, so he just stops paying his bill until he is disconnected). Additionally, the financial crisis in 2008 seems to have led to an anomalous increase in churn at the tail end of our dataset (and especially in the longitudinal holdout sample). These unforeseen macroeconomic forces may have influenced churn from all risks in some way, but the NPA risk was hit disproportionately. Therefore, it seems reasonable to classify these causes of churn into their own risk category. In our empirical application, we chose only three risks mainly for the sake of parsimony, but also because we lack any additional data that would allow us to make managerial recommendations based on a more granular classification of the risks. Our approach, though, could be generalized easily to accommodate reasons for churn that are categorized more coarsely or finely. For example, a company might want to differentiate between switching to a competitor and foregoing the service due to its price (both of which are combined in the Value risk).

One might be tempted to treat the 23 percent of customers for whom no cause of churn is reported as incomplete records, and omit them from the analysis. However, since churn is relatively rare in aggregate, let alone for any one particular risk or demographic category, we want to extract whatever information that we can from these customers. Dropping all of the incomplete records would be appropriate only if we assume that the data are both “missing at random” (MAR) and “observed at random” (OAR), as defined by Rubin (1976). OAR requires that “the conditional probability of the observed pattern of missing data, given the observed data and missing data, is the same for all possible values of the observed data” (Rubin 1976, p. 582). This assumption cannot hold, because it suggests that customers with different lifetimes, or from different geo-demographic clusters, are all equally likely to have their reasons for churn be recorded. But one could easily envision the willingness of a customer to report a cause of churn to vary along these factors. However, the MAR assumption is tenable, because it requires that “the conditional probability of the observed pattern of missing data, given the missing data and the value of the observed data, is the same for all possible values of the missing data” (Rubin 1976, p. 582). An alternative assumption, “not missing at random,” (NMAR) would assume that the missing data process actually depends on the relative likelihoods of the reasons for churn. While this is certainly a plausible assumption, we believe that it is unlikely. Also, estimating such a model would require additional information that is unavailable to us anyway. We can handle missing data under the MAR assumption using multiple imputation (Little and

Rubin 2002, ch. 10); details are in Appendix A.4.

Some summary statistics of the data are provided in Table 1. There are 48,693 households in the population from which we randomly assigned 43,867 households to a calibration sample (from which we estimate population-level model parameters), and a 4,826 to a cross-sectional holdout sample (that we use to assess the appropriateness of the model and parameter estimates through posterior predictive checks). Survivors are those customers who remained active through June, 2008.

	Calibration	Holdout
Total	43867	4826
Survivors	32194	3542
Stated reason for churn		
Value	1022	132
Personal	2239	259
NPA	5681	612
Missing	2731	281

Table 1: Number of households in calibration and holdout samples, with status at the end of the 15-month observation period.

3.2 The Model

Our model is a Bayesian hierarchical competing risk model, using the data likelihood from Section 2 as a foundation. The general form of the individual-level data model is in Equation (3), and depends only on θ_i . Assuming conditional independence across individuals, the full data likelihood is the product of Equation (3), across all customers $i = 1 \dots N$. Let r be the number of elements in θ_{ij} , so θ_i is the rJ -dimensional column vector that concatenates the J θ_{ij} 's for i . We now want to allow for heterogeneity in θ_i since some customers may be more likely to churn than others, and for different reasons. We also want to allow for correlation among elements of θ_i across the population, since some customers who are more likely to churn for one reason might also be more (or less) likely to churn for another. Furthermore, differences among customers could be explained in part by observable characteristics (e.g., geo-demographics), and in part by unobservable traits. Observed heterogeneity is captured through x_i , the p -length vector consisting of an intercept plus the $p - 1$ variables constructed from the geo-demographic indicator variables. Let Δ be the $rJ \times p$ matrix of coefficients on the log-linear regression

$$\log(\theta_i) = \Delta x_i + \varepsilon_i, \quad \varepsilon_i \sim MVN(0, \Sigma) \quad (6)$$

Our choice of functional form for $F_j(t_{ij}|\cdot)$ is a two-parameter Weibull distribution, parameterized such that $\theta_{ij} = [m_{ij}, c_{ij}]$ and $F_j(t_{ij}|\theta_{ij}) = 1 - 2^{-\left(\frac{t_{ij}}{m_{ij}}\right)^{c_{ij}}}$.³ Under this parameterization, m_{ij} is customer i 's latent median "lifetime" attributable to risk j , and c_{ij} is a shape parameter that affects duration dependence (positive when $c_{ij} > 1$, negative when $c_{ij} < 1$, none when $c_{ij} = 1$, and always monotonic) and tail behavior. We parameterize the Weibull distribution in terms of the median lifetime, as it allows for a more intuitive interpretation of coefficients when m_{ij} is regressed on covariates, and it is more computationally efficient than focusing on the mean of the distribution (which depends inconveniently on gamma functions). The Weibull reduces to the exponential distribution when $c_{ij} = 1$. While the exponential distribution and its discrete-time analog the geometric distribution have been employed in previous analyses of customer churn, the exponential distribution imposes a significant constraint in a competing risk model. Due to the constant hazard rate, resulting in the well-known "memoryless" property, the relative propensities of the different risks occurring are assumed to remain constant over the course of a customer's relationship. As we see in Section 3.3, such an assumption would not reflect the time-varying distribution of the reasons for churn that we observe in our data. Also, note that m_{ij} is not the same as the median duration of customer i 's relationship with the firm. It is the median lifetime for a single risk if that risk were the only possible reason for churn. We would expect that customers with high m_{ij} would be less likely to churn *from risk j* early in the relationship, but that some customers might be more likely to churn early when an m_{ij} is lower for a different risk j' . As we demonstrated in Section 2, using the risk-specific distribution $F_j(\cdot)$ from all j , we can construct the survival probability and hazard rate aggregated across risks, as well as the likelihood of an individual's observation.

The parameters Σ and Δ are the only model parameters that require an explicit declaration of prior knowledge. Since we have so little prior knowledge about these parameters, our choices arise primarily out of a desire for computational convenience. Our prior on Δ is placed on $\text{vec } \Delta$ (the vec operator transforms a matrix into a vector with the matrix columns concatenated end-to-end), and is multivariate normal with mean Δ_0 (a vector of the same length as $\text{vec } \Delta$), and covariance matrix $\Omega \otimes \Sigma$, where Ω is a pre-specified $p \times p$ matrix. We choose the prior on Σ to be an inverse Wishart distribution with location parameter A (an $rJ \times rJ$ matrix) and ν degrees of freedom, scaled such that the priors on the elements of $\log(\theta_i)$ are uncorrelated and weakly informative.

³Note that t_{ij} is observed only if the time is uncensored *and* risk j is the observed j_i . But there is still a distribution for churn time for all risks, and t_{ij} represents that time for risk j .

3.3 Estimation and Model Validation

Details about the MCMC algorithm are in Appendix A. We ran multiple versions of the model by varying the number of risks (1 or 3), and whether we included geo-demographic information. Parameters were estimated using data from the 43,867 households in the calibration set, but censored after March 2008, allowing for an additional three-month longitudinal forecast period. We also experimented with including some time-varying macroeconomic covariates, but found that increasing the dimensionality of the parameter space in this way added a lot to the computational requirements of the model without improving model fit or predictive performance.

Our model validation exercises focus on two managerially relevant quantities: retention rates and risk proportions. The retention rate at time t is simply $S(t)/S(t-1)$, or the proportion of those customers who survived through time $t-1$ who are retained in time t . The risk proportion is the percentage of those customers who churned in time t that churned for each of the J risks. To assess the reasonableness of our model as an illustration of customer behavior, we compare the observed churn patterns against the posterior predictive distributions of those patterns (Rubin 1984; Gelman et al. 1996). First, we look at how well the models capture overall retention behavior, regardless of the reason for which customers churn, using the mean absolute percentage error (MAPE) of our posterior predictions of the probabilities with which customers retain service each month, against the actual retention rates. We do this for the calibration and holdout samples, for the 15-month calibration and 3-month forecast periods, and for both the single-risk and competing-risk models. The “single risk” models ignore the cause of churn, are equivalent to the stochastic duration-based customer retention models that have appeared previously in the literature, and offer a reasonable benchmark for model comparison on this dimension.

For the aggregate retention patterns, during the 15-month calibration period, we find little variation in posterior predictive performance across the alternative model specifications. MAPEs are approximately .1% in the monthly retention probabilities of the calibration and cross-sectional holdout samples during the 15-month calibration period, and about 1.1% during the three-month forecast period. Though the close fit between the single-risk and our proposed competing risk framework may seem discouraging at first glance, this is in fact not surprising. The differences in the reasons for which customers churn is just one of many characteristics that make customers in a heterogeneous population different from one another. As such, a duration model that incorporates unobserved heterogeneity across customers should predict aggregate churn rather well. Similarly, it comes as no surprise that the models without geo-demographic variables offer comparable performance to those that incorporate such variables.

Plots of the posterior predictive retention curves, with 95% highest posterior density (HPD) intervals,

are in Figure 1. This figure is cohort-by-cohort for the cohorts who signed up from January to July, 2007 (we excluded the remaining cohorts, for whom we observe less information, solely for space considerations). Figure 1 shows how well the competing risk model captures data patterns in an absolute sense across multiple cohorts. Clearly the model and estimates of the population level parameters Δ and Σ are reasonable for both the calibration and holdout samples, in that they allows us to accurately replicate the data that was observed previously. Forecasts for the three months after the censoring point are, as one would expect, less accurate, but still mostly within the HPD intervals of the forecasts, which are themselves quite narrow in the range of retention rates considered. Unfortunately, the financial crisis of 2008 led to an overall decrease in retention rates during the forecast period in a way that we could not predict completely, but the model still seems to do quite well.

To assess the performance of the competing risk model in capturing the distribution of reasons for churn each month, we compare it to a benchmark that assumes the distribution is given by the mean of the empirical distribution that is observed throughout the calibration period. Under this benchmark, the distribution of the reasons for churn is assumed to remain unchanged from one month to the next. The competing risk model allows for duration dependence in the cause-specific hazard rates, consequently allowing for the proportion of customers churning due to each cause to vary as the tenure of the customer relationship progresses. The mean absolute difference between the observed and expected proportions churning due to each risk, averaged across risks and months, is 10.4% under the benchmark for both the calibration and holdout samples, but 2.8% and 2.9%, for the calibration and cross-sectional holdout samples, respectively, for the posterior predictive mean of that distribution.

Figure 2 shows the actual, and the 95% HPD contours for the posterior predictive distributions, for the proportion of customers who, conditional on churning some months into the relationship, churn for each of the three possible risks. Any point within the triangle represents a trivariate vector of probabilities. The arrows along each axis indicate which set of gridlines correspond to that axis. Numbers within the triangle represent the elapsed time since acquisition. The location of the number within the triangle represents that *observed* trivariate vector of the proportions of customers in that cohort who churn due to a particular risk. The contour lines around each number indicate the 95% HPD region of the posterior predictive density for the probability vector. In the figure, we see that customers who churn early in their tenure are most likely to churn from the NPA risk, but that as time elapses, the NPA risk becomes less likely and the Personal risk becomes more so. Eventually, we see more customers churn for value-related reasons, but that risk is not as prevalent early in the relationship. This pattern is easily explained as a “sorting effect” (Fader and Hardie 2010) where as the NPA-prone customers churn out, their proportion of the population declines. The posterior predictive HPD contours track the actual churn proportions reasonably well. Note that given

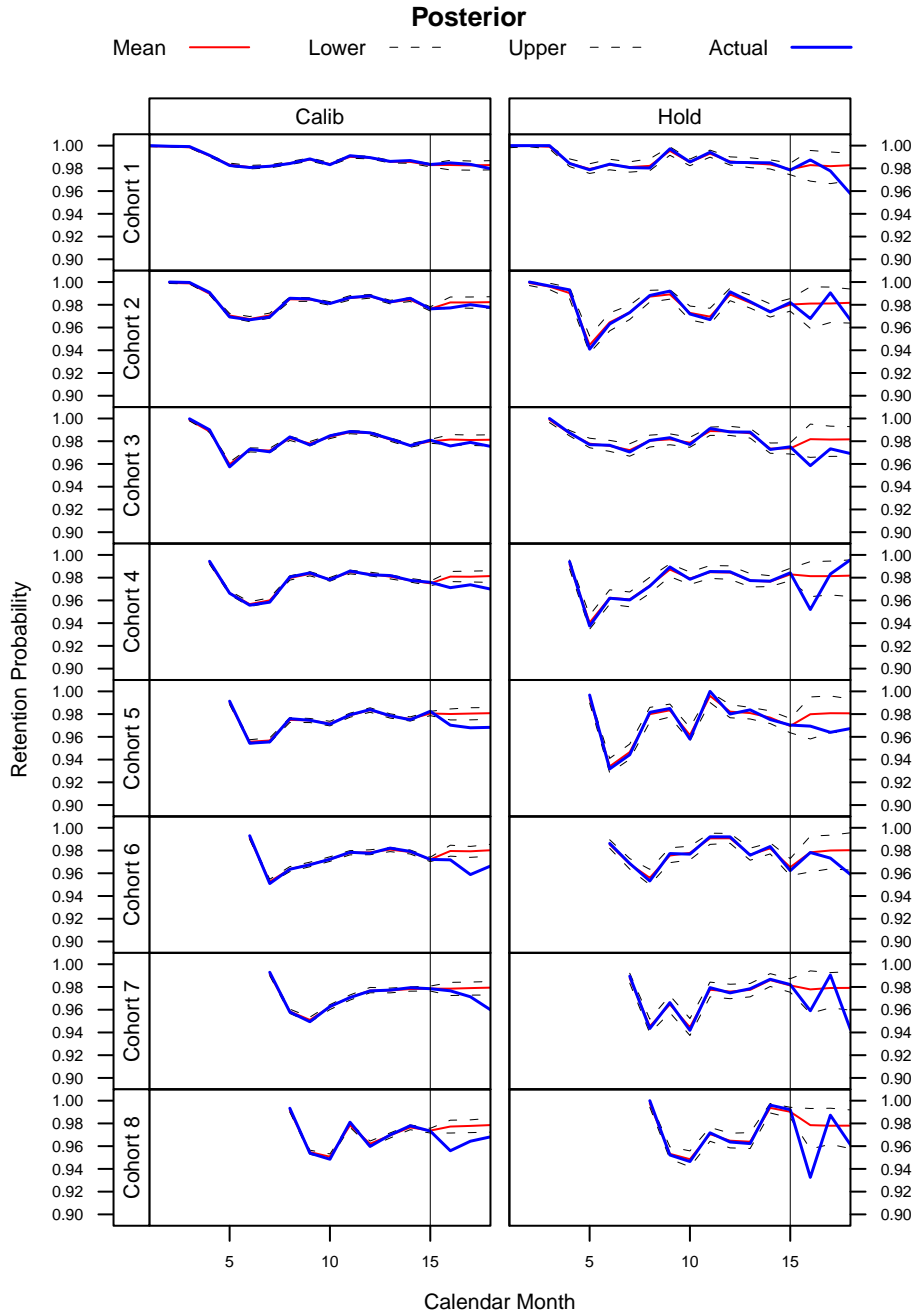


Figure 1: Mean, and 95% HPD intervals for posterior predictive retention curves, cohort-by-cohort

the smaller number of customers who are at risk deep into a relationship (i.e., only those in the early-2007 cohorts), both the data and predictions are noisier than those who churned earlier (there are just more of those people in the dataset to consider).

3.4 Interpretation of parameter estimates

Table 2 summarizes the posterior means and 95% HPD intervals of parameter estimates for the competing risk and single-risk models, with and without demographic information. The first two lines are the median times to churn from each of the three risks (the Value, Personal and NPA columns) from the competing risk model, and the aggregate median lifetimes (the All column) from the single-risk model. The reported median lifetimes under the model with demographics are for customers in a baseline demographic group (urban, low income homeowners with no children, some high school education and a blue-collar/service job). Note that the median risk-specific lifetimes, and the aggregate lifetimes, are lower for this baseline group than for the population at large, suggesting that these customers are more likely to churn sooner. We also report the percentage difference in risk-specific lifetimes for different levels of each of the demographic factors. For example, the posterior mean of the median time to churn because of the Value risk is 16.7% longer for high income customers than for low income customers.

Table 2 reveals a critical point: the one risk of the three that is thought to be controllable by the firm is the risk that stimulates churn least often. Also, there are substantial differences across geo-demographic characteristics for each of the different reasons for churn that we consider. One interesting pattern is that the signs of the percentage differences tend to be the same from risk to risk, suggesting that “ordering” of which risks are most prevalent in each cluster remains the same from cluster to cluster. It is interesting to note that we generally observe the largest differences in churn propensities due to Non-Pay/Abuse (NPA), which also happens to be the most prevalent cause of churn for all groups.

We can see these variations more clearly in Figure 3. Each panel represents the posterior means of the hazard rate functions for 24 months, broken down by geo-demographic characteristic factor (rows) and level of those factors (columns). Each curve in the panel is either risk-specific hazard rate (the broken lines) or the aggregate hazard rate function (solid line). The panels on the right of the page are the hazard rate functions for the entire population, which is replicated across rows to facilitate comparison. This geo-demographic breakdown illustrates the canceling-out effect that can occur when modeling unobserved heterogeneity. Not only does the aggregate hazard rate function vary across groups, but the differences among risk-specific hazards varies as well. In particular, note that the hazard for Non-Pay/Abuse is relatively high at the start of a customer’s relationship, but declines sharply as time passes. These differences,

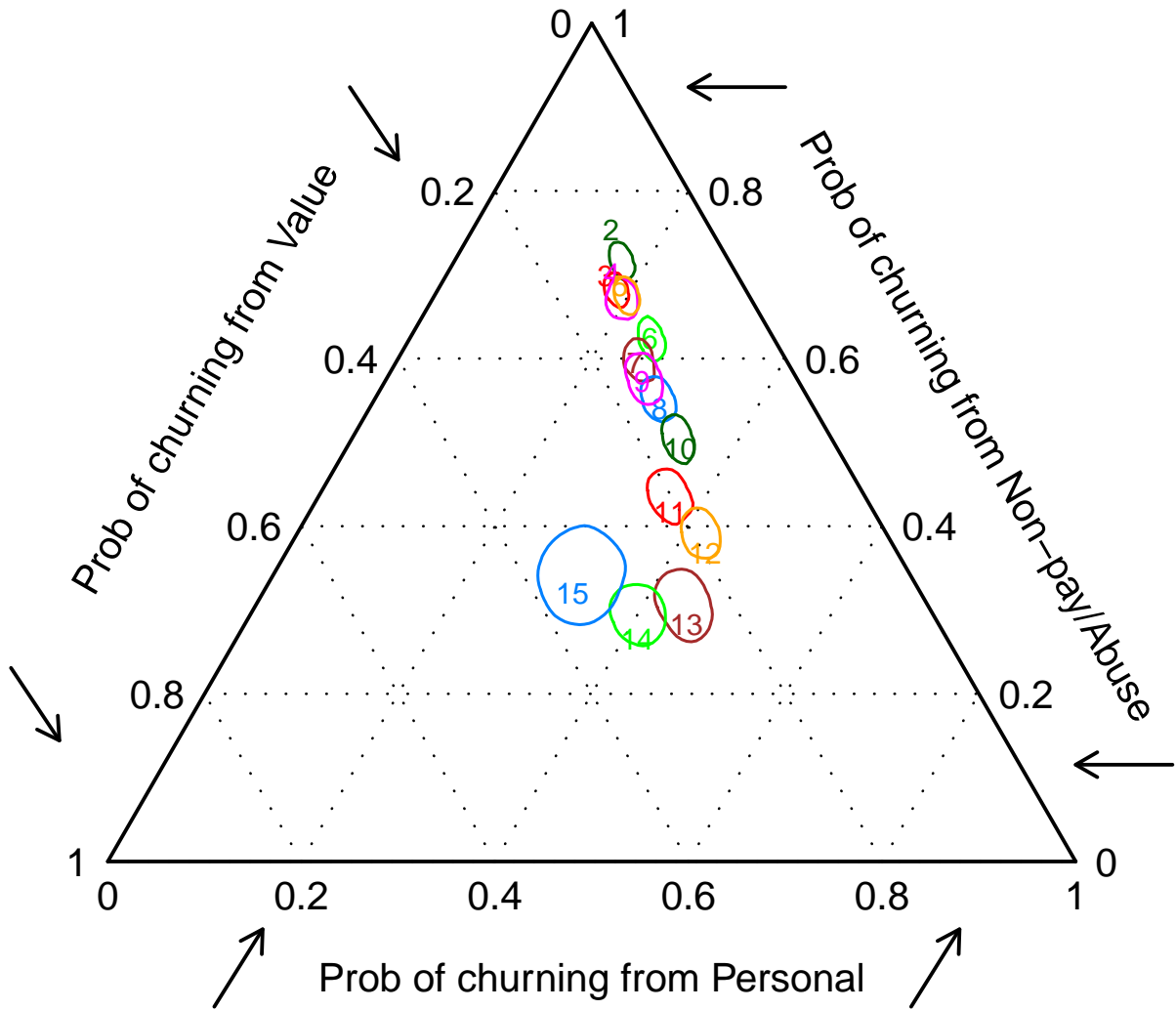


Figure 2: Observed and posterior HPD contours for the relative proportions of reasons for churn for customers who have survived a certain number of months. Any point within the triangle represents a trivariate vector of probabilities. The arrows along each axis indicate which set of gridlines correspond to that axis. Numbers within the triangle represent the elapsed time since acquisition. The location of the number within the triangle represents that *observed* trivariate vector of the proportions of customers in that cohort who churn due to a particular risk. The contour lines around each number indicate the 95% HPD region of the posterior predictive density for the probability vector.

	Value		Personal		NPA		All	
	HPD Lower	HPD Upper	HPD Lower	HPD Upper	HPD Lower	HPD Upper	HPD Lower	HPD Upper
Models with no demographics								
Median lifetime	103.9	140.0	57.3	65.2	45.1	49.7	35.1	38.0
Models with demographics								
Baseline median lifetime (months)	92.3	121.4	52.7	60.3	40.1	43.9	31.5	34.0
<i>Percentage differences between groups</i>								
Loc - Town/Rural : Urban	0.6	0.9	0.8	0.9	1.1	1.2	0.9	1.1
Loc - Suburban : Urban	-6.5	-3.7	-0.8	3.5	-5.1	-1.4	-3.6	-0.4
Income - Middle : Low	-12.6	-10.7	-14.6	-12.3	-16.8	-14.9	-14.8	-13.1
Income - High : Low	13.6	16.7	18.2	22.0	23.8	27.5	20.4	23.6
Age - Middle : Low	9.7	12.2	8.2	12.1	14.9	18.6	12.0	15.3
Age - High : Low	-4.6	-3.2	-2.2	-0.2	-4.5	-2.8	-3.7	-2.1
Kids - Yes : No	-0.2	2.7	-4.6	-0.5	-0.4	3.5	-1.1	2.2
Home - Rent : Own	-12.2	-9.3	-16.5	-12.9	-16.4	-13.2	-15.0	-12.1
Job - White Collar : Blue Collar	0.5	0.6	0.4	0.6	0.7	0.9	0.6	0.7
Job - Prof/Mgmt : Blue Collar	8.7	11.3	14.2	17.8	16.6	19.8	14.9	17.6
Job - Retired : Blue Collar	-3.4	-2.5	-1.7	-0.3	-3.5	-2.2	-2.9	-1.7
Edu - HS Grad : Some HS	-6.2	-5.2	-7.5	-6.2	-8.5	-7.4	-7.5	-6.5
Edu - Some Coll : Some HS	-9.4	-8.0	-8.5	-6.5	-11.6	-10.0	-10.0	-8.5
Edu - Coll Grad : Some HS	7.5	9.4	6.4	9.4	11.5	14.3	9.3	11.7
Edu - Grad Plus : Some HS	5.0	6.6	8.7	11.0	9.7	11.8	8.7	10.5

Table 2: Baseline and percentage differences in median lifetimes, for each risk and all risks in aggregate, for groups with different demographic characteristics. For example, the posterior mean of the median time to churn because of the *value* reason is 16.7% longer for high income customers than for low income customers. Lower and upper bounds represent the 95% highest posterior density (HPD) intervals. The baseline lifetime in the first row represents customers who are urban, low income, homeowners with no kids, “blue collar” employment and some high school education.

both across geo-demographic clusters and over time, will play a large part in our analysis in the next section, in which we show how altering churn propensities for different risks, for different geo-demographic clusters, and at different points in the relationship can have varying effects in the long-term value of the customer base.

4 Implications for Customer Retention and Lifetime Value

Financial metrics that are derived from measures of service retention, such as customers' residual value (Fader and Hardie 2010) and expected lifetime value (Schweidel et al. 2008b), remain important tools in marketing, as managers look for ways to better understand in which of their customers they should invest and which are expected to generate the greatest profits. With such information, managers can allocate resources across the customer base more efficiently. In this section, we demonstrate how the competing risk model provides new insights into customer retention and value patterns that are not afforded by models that do not consider the reason for which customers cancel service.

With parameter estimates from our selected models in hand, we now turn to important questions of how managers can use insights from the competing risk framework to manage their businesses. Recall that in our empirical application, we consider three different possible causes of churn. We assume that one, Value, is "controllable" by the firm in the sense that it could use actions such as marketing, pricing, and service improvement to potentially delay the likelihood that a customer would churn due to this risk. For now, let's assume that the other risks are uncontrollable. The Personal risk is recorded for customers who move from the service area, die, or otherwise cancel for reasons completely outside the control of the firm. The Non-Pay/Abuse risk may or may not be "controllable" by the firm. For illustrative purposes, we initially assume that it is uncontrollable and then relax this assumption.

There are two ways that we can think about the lifetime value of the customer. One is expected customer lifetime value (ECLV), which we define as the stream of future expected cash flows, discounted back to the present. Let θ_{ij} be a heterogeneous parameter of the risk-specific churn process for risk j (such as the median time-to-churn, as in our Weibull model), and let θ_i be the vector of all J of these parameters. Assuming a \$1-per-time-period future revenue stream from a single customer, and a known, constant and homogeneous discount rate δ , then

$$ECLV(\theta_i) = \sum_{t=1}^{\infty} S(t|\theta_i) \delta^t \quad (7)$$

As the survival probability decreases over time, it becomes less and less likely that the firm would ever

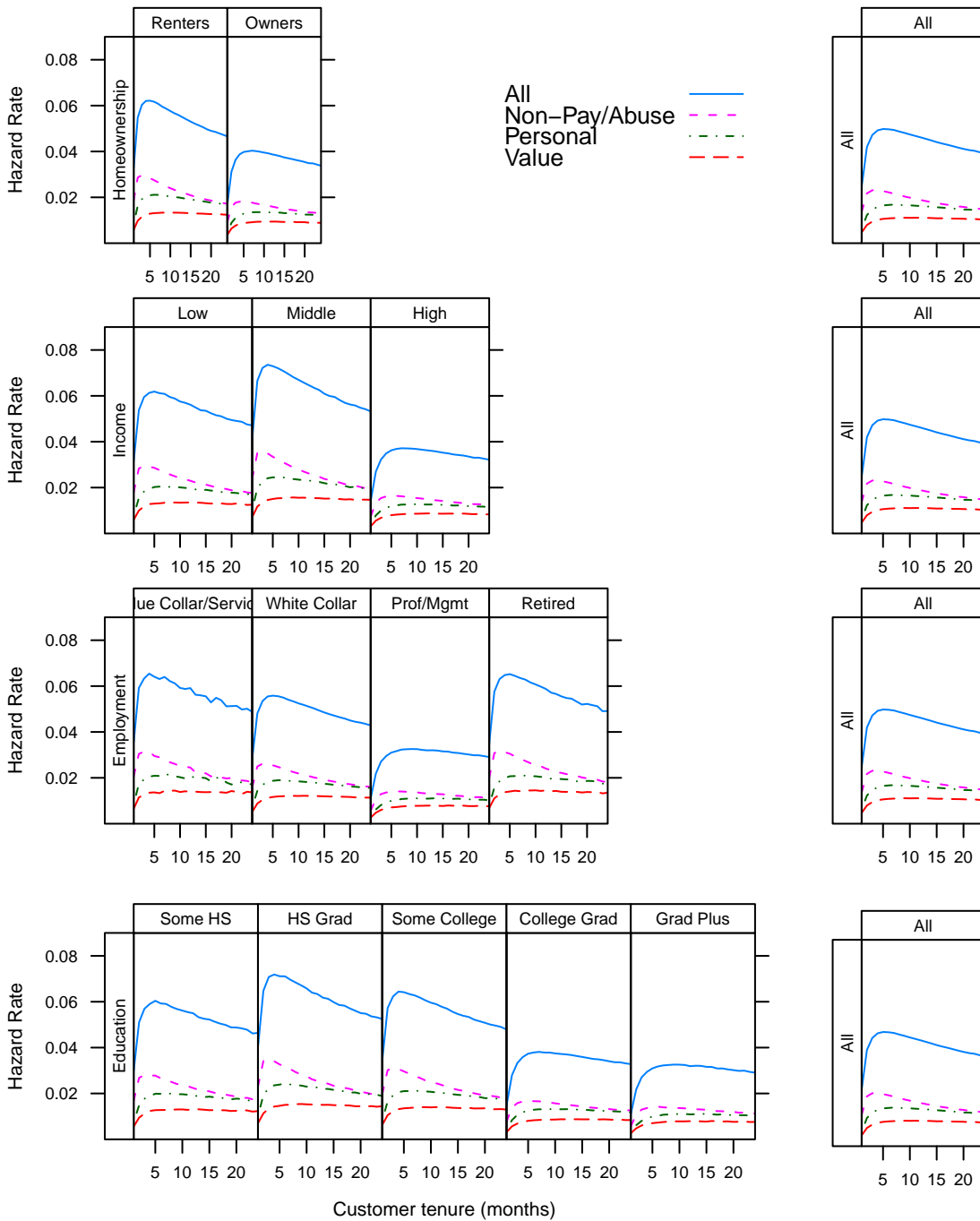


Figure 3: Posterior mean hazard rate functions, by demographic group, for each of the competing risks, and all risks combined. Panels in each row correspond to a demographic factor (Homeownership, Income Level, Employment Type and Education Level). The “All” panels on the right correspond to the entire population, and are replicated across rows to facilitate comparisons.

realize the discounted revenue from future periods. Naturally, if a firm were to “slow down” the churn rate of its customers, ECLV would increase, as customers would remain with the firm longer. When we consider multiple risks, any of which can cause a customer to terminate service, not all of them are within the control of the firm. What the manager needs to know is the incremental effect of delaying churn due to the risks which they can influence, taking into account the fact that any of the other risks may also trigger customer churn.

In Appendix B, we show that the marginal effect of a change in θ_{ij} on ECLV is

$$\frac{\partial ECLV(\theta_i)}{\partial \theta_{ij}} = \sum_{t=1}^{\infty} \delta^t \frac{S(t|\theta_i)}{S_j(t|\theta_{ij})} \frac{\partial S_j(t|\theta_{ij})}{\partial \theta_{ij}} \quad (8)$$

The factor on the right in the summand in Equation (8) is how much the probability that customer i survives risk j until time t changes per unit change in θ_{ij} , holding all of the other $J - 1$ elements of θ_i constant. If θ_{ij} is the median time-to-churn from risk j , then this marginal effect factor is positive. If we are only considering a single risk, then $S_j(t|\theta_{ij})$ is equal to $S(t|\theta_i)$, and the middle factor, a ratio between the aggregate and risk-specific survival probabilities, is 1. All of the discounted “improvement” in the survival probability manifests in ECLV. When a customer faces multiple risks, however, the ratio of $S(t|\theta_{ij})$ to $S_j(t|\theta_i)$ is less than 1. This ratio acts as a *dampener* on the effect of efforts to delay churn because of risk j . Intuitively, if risk j is a prevalent cause of churn in the population, then much of the effect of delaying churn from risk j still passes through as an increase ECLV. But if churn due to risk j is rare, and it is likely that customers churn for reasons other than risk j , then reducing churn due to risk j won’t have much of an effect on ECLV at all.

This preceding analysis of ECLV is helpful for illustrating how the interplay among risk-specific churn processes influences ECLV. But in reality, we don’t know what θ_i is for a particular person. We do, however, know how long it has been since a customer was acquired, and we can use this information to update our beliefs about any existing customer’s own propensity to churn. Fader and Hardie (2010) introduce this idea when discussing discounted expected residual lifetime (DERL). Like ECLV, DERL is a discounted stream of expected future cash flows, but for a customer who was acquired T_i periods ago. Given θ_i ,

$$DERL(T, \theta_i) = \sum_{t=T}^{\infty} S(t|\theta_i, t > T) \delta^{t-T+1} \quad (9)$$

To account for variation in θ_i , we need to take the expectation of Equation (9) with respect to the posterior distribution of θ_i , given that the customer has survived T_i periods already. Fader and Hardie (2010) do this in the single-risk case, under different distributional assumptions and without demographic information,

and are able to get an analytical result for the posterior expected DERL. They also derive an analytical expression for “elasticity of retention”, quantifying the connection between retention rates and expected lifetime value, and illustrating that the sensitivity of DERL to retention rates varies according to the patterns of heterogeneity in the population. For a hierarchical competing risk model, we can numerically intergrate over the posterior distributions of θ_i (using importance sampling to generate posterior predictive draws of θ_i) and estimate a posterior expected DERL for each person. Figure 4 plots how DERL, discounted back to today, for a hypothetical customer in each of five geo-demographic groups, would change for different durations of service T . Descriptions of these clusters are in Table 3. Naturally, customers who have a longer tenure with the firm are less likely to have high churn rates, so DERL always increases with T . Note that in this illustrative exercise we are projecting the model deep into the future, maintaining parameter estimates that were generated from data collected during a narrow observation window. Of course, the state of the world can change in the future, but these estimates of DERL assume future stationarity of the parameters.

Cluster	Urbanicity	Income	Age	Kids	Own/Rent	Employ	Edu
A	Urban	High	Middle	Yes	Rent	White Collar	College Grad
B	Town/Rural	High	Middle	Yes	Own	Prof/Mgmt	College Grad
C	Suburb	Middle	High	No	Own	Retired	Some College
D	Suburb	High	Middle	No	Own	Prof/Mgmt	Grad Plus
E	Urban	Middle	Low	Yes	Rent	White Collar	Some College

Table 3: Characteristics of five selected demographic groups.

One remaining question, of high salience for managers, is what the impact of firm activity will be on DERL. Equation (8) shows that the impact of a firm’s actions on ECLV for a new customer depends on the relative prevalence of these risks, and this same insight applies to DERL as well. What makes answering this question for DERL more complicated is that we now need to appreciate that the passage of time provides some information about the a customer’s latent propensensity to churn from each one of the possible risks. To untangle these factors, we consider the following hypothetical situation. Suppose, for customers in a particular geo-demographic group, a manager can choose to pull any of J levers, and can decide how far to pull that lever. Pulling a lever is analogous to investing in efforts that are intended to retain customers longer (e.g., price promotions, service improvements, advertising, or even bill collection). Depending on how far the manager pulls lever J , he delays the median time of churn due to risk j by μ_j additional years. Put another way, recall that the time for a customer to churn from risk j is a random variable, with some risk-specific distribution. We allow the manager to change that distribution in such a way that the median *remaining* time to churn is extended by μ_j . The mechanism behind this adjustment is described in Appendix C. By changing the remaining median lifetime by μ_j , he changes DERL from the baseline level in Equation (9) to something higher. But this difference in DERLs depends on θ_i , which is

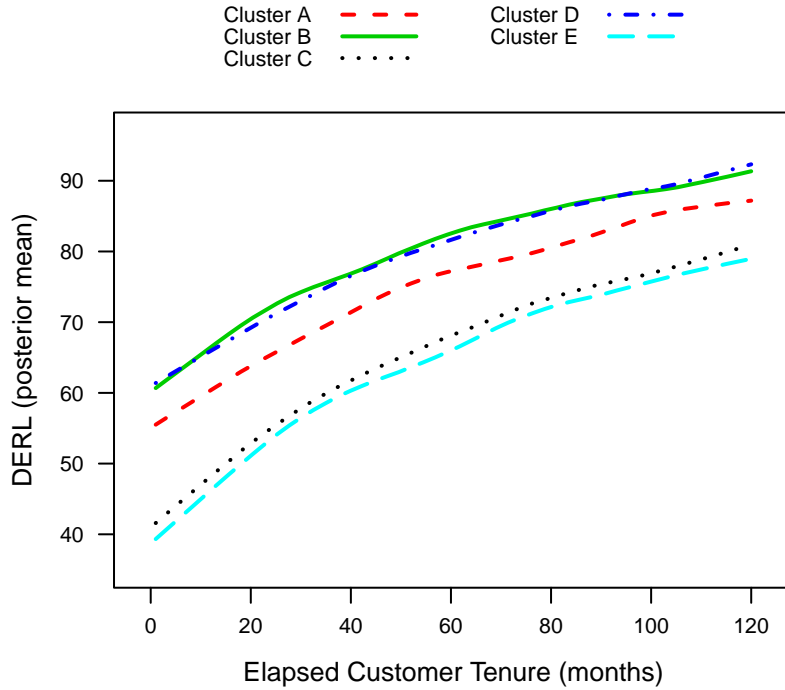


Figure 4: Posterior expected values for discounted expected residual lifetime (DERL) for five geo-demographic clusters. Churn propensities for customers with long tenure with the company are more likely to be lower than more recent customers, and so we expect their lifetimes to be longer. The descriptions of the clusters are in Table 3

unobserved. Integrating the difference between the baseline DERL and the “ μ_j -enhanced” DERL, over the posterior distribution of θ_i given T , gives the manager an upper bound on the amount he should “spend” to pull lever j . With θ_i integrated out, DERL depends on T , which is something the manager does observe directly, and μ_j , which is something he can influence indirectly through his managerial actions (pulling the levers). We refer to the expected difference between the baseline DERL and the “ μ_j -enhanced DERL” as the expected incremental DERL.

Figure 5 illustrates the relationships among T , μ_j and expected incremental DERL, for customers in the same five geo-demographic clusters that were described in Table 3 and used for Figure 4. Each row of panels is a cluster, and each column of panels represents one of the J risks. The x-axis is μ_j , and the y-axis is T . These plots show how much additional DERL the firm can expect to get by adding μ_j to risk J 's lifetime, for a customer who is T years since acquisition. Immediately, we see that this incremental benefit depends on which risk the firm is trying to manipulate. No matter how long the customer has been with the firm, delaying churn that is due to Value-related reasons, has relatively little affect on DERL. Keeping Equation (8) and Figure 3 in mind, this should not be very surprising. Since Value is the least prevalent

risk in the population, customers are more likely to churn from something else first (the dampening factor is closer to zero). But this is a sobering realization for a firm for which Value is the only risk over which it has any influence. Depending on the how much it costs to add to the Value-specific remaining lifetimes, and how much each unit of DERL is worth to the firm, the firm could find itself accruing very little return for its retention efforts.

For each of the reasons for which customers churn, we observe backward bending patterns in the surfaces in Figure 5. This is a consequence of two forces. First, a customer who has remained with the firm for a long time will have a lower *posterior* expected churn propensity than newer customers, as they would not have survived T years otherwise. As such, those customers who have already survived for a long duration T have long expected remaining lifetimes, as illustrated in Figure 4. These findings are consistent with a sorting effect in the presence of positive duration dependence (Fader et al. 2009), as those customers who are most prone to churn do so early, resulting in the initial dip in retention seen in Figure 1, leaving only those individuals who are less prone to churn.

Second, the impact of adding *additional* months to the remaining median lifetime depends on the extent of discounting. Consider two customers with the same expected latent lifetimes, dictated by θ_{ij} , but who differ in their tenure with the service provider. The customer with a longer tenure is closer to the end of his latent lifetime. As such, these additional months are not discounted as heavily for him compared to the younger customer for which the effect of discounting is greater. After taking the discounting into account, adding additional months that are far into the future may do little to influence DERL.

The net of these two effects varies across the geo-demographic clusters. There is very little difference across clusters when considering delaying Value-related churn. DERL for Cluster D is *slightly* more sensitive to μ_j in the Value risk than the other clusters, mainly because Value-related churn is more prevalent in the suburbs than in other places (see Table 2). But the incremental DERL “surface” is low and flat for this risk, across the population, compared to the other risks. No matter how much extra μ_j the firm adds to a customer, it only recovers 1-3 additional DERL units. Given a choice, a firm might not want to invest in delaying Value-related churn for anyone. If it were to do so, however, such efforts are best spent on “older” customers.

While Value, the only “for sure” controllable risk in our empirical application, is the least prevalent risk and hence the one for which changes to the median remaining lifetime will have little effect on DERL, this may be idiosyncratic to our data provider. For other firms, the controllable risk might be the most prevalent one, or may be in the middle. Suppose, for example, that the firm could take actions to reduce the extent of churn due to the Non-Pay/Abuse risk. As we see from Figure 5, doing so would be quite fruitful. That, again, should not be surprising in light of Equation (8). If you reduce the causes of churn

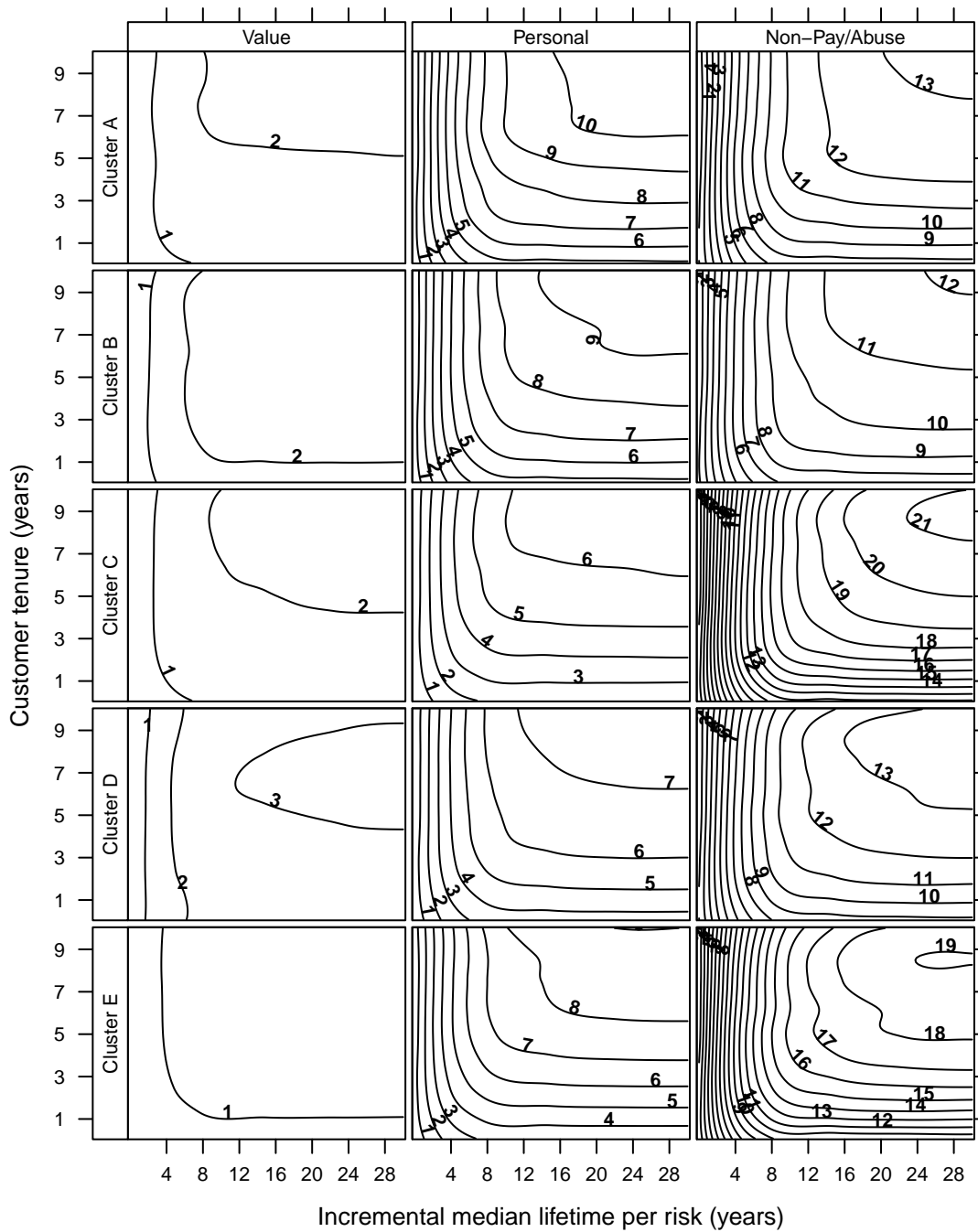


Figure 5: Contour lines representing the incremental discounted expected residual lifetime (DERL) that a firm can expect by delaying churn that is due to a specific risk. The x-axis represents the additional number of years added to the remaining median risk-specific lifetime for each risk. The y-axis represents the number of years of the tenure of the existing customer relationship. DERL was computed using an annual discount rate of 8 percent.

that are more likely to actually generate churn, then more customers will be retained longer. What Figure 5 offers the practicing manager are numbers that can be estimated from historical data that the firm already possesses.

Hypothetically, suppose that the firm could in some way influence the “Personal” risk. We observe no greater variation across the geo-demographic clusters. Cluster A consists of urban renters with children, and so one should not be surprised if these customers are more prone to move out of the service area. Delaying relocation from those customers will have more of an effect on DERL than doing it for cluster C, which is comprised of older and retired suburbanites. For cluster C, churn for Personal reasons is not zero, but there are other reasons for churn that dominate in this cluster. Also, for this risk we can see how delaying churn might have different effects on the expected return on DERL, depending on the elapsed tenure of the customer. Customers who are farther along in their lifetimes benefit more from longer extensions to their “end dates.”

Finally, what if the firm could affect churn due to Non-Pay/Abuse? For some clusters like C and E, there is a higher *prior* belief that members will churn from this risk quickly. When a customer in one of these clusters survives for a long time, it becomes more and more likely that the median lifetime of this customer, for that risk, is in the right tail of the distribution. That’s why the NPA contour lines for those clusters start to turn back on themselves. These customers might be farther from the start of their relationship with the firm, but since we infer a very long lifetime, they may still be very far from the end.

5 Discussion

Our hierarchical competing risk model jointly captures “lifetime” for customers of a contractual service provider and the reasons behind the cancellation. The analysis reveals that while single-risk duration models are sufficient for modeling the time until customers churn, the competing risk framework is required when managers care about heterogeneous patterns in causes of churn. In the presence of multiple causes of churn, the economic return on retention management activities depends on variation in customer characteristics and customer tenure with the firm, as well as the specific tactics that the firm chooses to deploy. Competing risk models are established, conceptually simple, and straightforward to estimate using existing methods of Bayesian inference. This article describes a novel approach to applying competing risk models to an important managerial setting and offers insights that for managers that extant retention models do not.

As we illustrate, disentangling the likelihood of churn due to different reasons provides managers with a clear understanding of the potential impact that marketing actions (Rust et al. 2004) and service

improvements (Rust et al. 1995) can have on customer churn. The effect of slowing churn due to particular causes varies during the course of a customer's relationship with the firm because of temporal variation in the strength of the dampening effect from alternative causes of churn (Equation 8). As the relative likelihood of churning due to each risk varies with a customer's tenure, so too will the impact of marketing actions, suggesting that the firm's retention strategy be dynamic (Hogan et al. 2002). The information provided by the competing risk model can then be incorporated into resource allocation decisions, enabling the service provider to invest in those customers for which it is expected to provide the largest return (Venkatesan and Kumar 2004). As customers' remaining values change over time, this may result in the firm focusing its efforts on different customers at different times. To facilitate managerial decisions, measures derived from the competing risk model, such as the incremental DERL, can be incorporated into marketing dashboards (O'Sullivan and Abela 2007).

In many respects, our research serves as a caveat that managers should not necessarily focus their efforts on raw retention metrics. Much like claims that 100% loyalty was an important managerial objective (Reichheld and Sasser 1990; Jones and Sasser 1995) have been countered because they are unobtainable (Oliver 1999), recent work has urged managers to be more thoughtful in their customer retention tactics (Keiningham et al. 2005; Sharp 2010). Our results are consistent with this view and provide a clear illustration of the obstacles that firms face when managing customer retention.

There are a number of directions in which the current research could be extended. Incorporating time-varying covariates is conceptually straightforward, though it increases the computational burden of the analysis. Early in our empirical investigations, we considered incorporating a measures of consumer sentiment and other macroeconomic variables, but did not find that it contributed substantively to our analysis (this may, in part, be attributable to the unforeseeable economic upheaval in 2008, but in practice, the model fits extremely well as it is). Other contexts may warrant the inclusion of covariates such as marketing actions or any touchpoints the service provider has had with the customer (Seetharaman and Chintagunta 2003). In doing so, it would be prudent for marketers and researchers to consider the potential for endogeneity and strategic thinking on the part of customers. For example, a customer may submit a number of complaints, anticipating that the service provider will respond by offering a financial discount. We believe that incorporating such aspects into models for customer base analyses is an important direction for research, as it poses both methodological challenges and has considerable practical value.

It may also be worth reexamining the way in which firms report retention to their stakeholders. While an aggregate measure of customer retention can provide one assessment of financial health, it is interesting to consider how stakeholders may respond to reports of the effectiveness of the firm's marketing activities as contributing to customer retention (Lehmann 2004). Should firms decide to report measures of

marketing's impact on increasing retention, it may be in their interests to provide metrics that recognize the dampening effect that we have discussed, allowing stakeholders to more clearly evaluate the impact of the service providers investments in marketing and service improvements. The same holds true within an organization, as marketers need to demonstrate the efficacy of their investments. In presenting such metrics, future work may also consider how the way in which the information is displayed influences how it is evaluated (Raghubir and Das 2010).

While an extreme interpretation of our results might suggest that we advise firms to halt their investments in activities that would delay customer churn, this would not be an appropriate inference to draw from our work. The return on the firm's investment in retention depends on both the prevalence of the different risks in play, as well as the cost associated with its retention efforts. Our intent is to provide managers with a tool they can use to help organize all available information when their decisions. Also, for the purposes of our illustration and to keep the analysis tractible, we assumed that the marketing "levers" affect only a single cause of churn. If a firm's efforts simultaneously influence multiple risks, though potentially to different extents, the incremental benefit of their efforts will be greater, which should be taken into account when comparing the expected benefit to the cost of the retention activity.

For a firm to maintain its market share, it will have to replace the customers it loses (whether from controllable or uncontrollable churn) with new ones (Ehrenberg et al. 1990; Sharp 2010). While our focus in this research has been on the incremental value of retention, the firm's retention efforts may also influence the acquisition process. As marketing actions and service improvements may simultaneously influence both the acquisition and retention processes, understanding the full impact of the firm's actions necessitates that these processes be considered jointly (Schweidel et al. 2008a). With a fixed budget to be allocate across prospects and existing customers of different tenures, doing so could provide guidance as far as how resources should be deployed. Given that we find that the reasons for which customers are likely to churn change during the course of their tenure, an "optimal" balance between acquisition and retention activities may depend on current customers' tenures to date (Reinartz et al. 2005). Doing so may reveal that marketing efforts and improvements to service quality, for example, play a much larger role in the acquisition process, thereby helping to replenish the customer base from those customers who have churned due to uncontrollable reasons. It may also be worthwhile to see if a relationship exists between the set of services to which customers subscribe at multi-service providers or their usage behavior and their eventual reason for discarding service. As the other drivers of customer value are examined, it remains essential that subsequent analysis ultimately considers how such relationships may impact the firm's bottom line, maintaining the managerial relevance of the research.

Appendices

A MCMC Algorithm for Model Estimation

We use Gibbs sampling to estimate marginal posterior distributions of $\theta_i, i = 1 \dots N$, Δ and Σ . To clarify some of this exposition, we need to introduce some additional notation. Let Θ be the $N \times rJ$ matrix where each row is θ_i' (e.g., the parameters of the J risk-specific Weibull timing distributions for household i). and let X be the $N \times p$ matrix where each row is x_i' (the vector of covariates for person i , including an intercept). The symbol $A \otimes B$ is the Kronecker product of A and B . The conditional posterior distributions are as follows:

A.1 Sampling θ_i .

Under the assumption of conditional independence, we can sample θ_i for each customer sequentially. The prior for each θ_i is

$$\pi(\theta_i|\Delta, x_i, \Sigma) \propto |\Sigma|^{-\frac{1}{2}} \exp \left[-\frac{1}{2}(\theta_i - \Delta x_i)' \Sigma^{-1} (\theta_i - \Delta x_i) \right] \quad (10)$$

The general form of the log conditional posterior (ignoring normalizing constants) is the log of data likelihood in Equation (3), plus the log of the prior distribution in Equation (10).

$$\log \pi(\theta_i|t_i, j_i, \Delta, x_i, \Sigma) = \log \pi(\theta_i|\Delta, x_i, \Sigma) + \begin{cases} \log [S_j(t_i - 1|\theta_{ij}) - S_j(t_i|\theta_{ij})] + \log S(t_i - 1|\theta_i) & \text{if } d_i = 1 \\ \log S(T_i|\theta_i) & \text{if } d_i = 0 \end{cases} \quad (11)$$

Under the median-parameterized Weibull distributions for the risk-specific churn models, $\theta_{ij} = [m_{ij}, c_{ij}]$ the risk-specific survival function is

$$S_j(t_i|\theta_{ij}) = 2^{-\left(\frac{t_i}{m_{ij}}\right)^{c_{ij}}}, \quad (12)$$

from which we get $S(t_i|\theta_i)$ using Equation (5). For the censored case, replace t_i with T_i .

The unnormalized log posterior distribution in Equation (11) is not a standard form, but there are numerous methods that one can use to simulate θ_i from it. For any multiple-risk model, θ_i will be of sufficiently high dimension that some form of adaptive Metropolis-Hastings algorithm is a reasonable way to go.

A.2 Sampling Δ .

The prior distribution on $\text{vec } \Delta$ is multivariate normal, with mean Δ_0 and covariance $\Omega \otimes \Sigma$.

$$\pi(\text{vec } \Delta | \Delta_0, \Omega, \Sigma, x) \propto |\Omega \otimes \Sigma|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \text{vec}(\Delta - \Delta_0)' (\Omega^{-1} \otimes \Sigma^{-1}) \text{vec}(\Delta - \Delta_0) \right] \quad (13)$$

The conditional posterior of Δ also depends on the prior on Θ , which we get by multiplying the priors of all θ_i .

$$\pi(\Theta | X, \Delta, \Sigma) \propto |\Sigma|^{-\frac{N}{2}} \exp \left[-\frac{1}{2} \sum_{i=1}^N (\theta_i - \Delta x_i)' \Sigma^{-1} (\theta_i - \Delta x_i) \right] \quad (14)$$

Equation (14) depends on Δ , not $\text{vec } \Delta$. So, expressing Δx_i as $I_{rJ} \Delta x_i$ (I_{rJ} is the $rJ \times rJ$ identity matrix), and applying the identity $\text{vec}(ABC) = ((C' \otimes A) \text{vec } B)$, we write the joint prior distribution for Θ as

$$\pi(\Theta | X, \Delta, \Sigma) \propto |\Sigma|^{-\frac{N}{2}} \exp \left[-\frac{1}{2} \sum_{i=1}^N (\theta_i - (x_i' \otimes I_{m(k+g)}) \text{vec } \Delta)' \Sigma^{-1} (\theta_i - (x_i' \otimes I_{m(k+g)}) \text{vec } \Delta) \right] \quad (15)$$

No other terms in the joint posterior distribution involve Δ , so we can get the conditional posterior distribution for $\text{vec } \Delta$ by multiplying Equations (13) and (15) together, and simplifying the result by ‘‘completing the square.’’

$$\begin{aligned} \pi(\text{vec } \Delta | \theta, \Omega, \Sigma, \Delta_0) &\propto \exp \left[-\frac{1}{2} \text{vec}(\Delta - \Delta_0)' (\Omega^{-1} \otimes \Sigma^{-1}) \text{vec}(\Delta - \Delta_0) \right. \\ &\quad \left. + \sum_{i=1}^N (\theta_i - (x_i' \otimes I_{m(k+g)}) \text{vec } \Delta)' \Sigma^{-1} (\theta_i - (x_i' \otimes I_{m(k+g)}) \text{vec } \Delta) \right] \\ &= \exp \left[-\frac{1}{2} \left(\text{vec } \Delta' \left((\Omega^{-1} \otimes \Sigma^{-1}) + \sum_{i=1}^N (x_i \otimes I_{m(k+g)}) \Sigma^{-1} (x_i' \otimes I_{m(k+g)}) \right) \text{vec } \Delta \right. \right. \\ &\quad \left. \left. - 2 \text{vec } \Delta' \left((\Omega^{-1} \otimes \Sigma^{-1}) \text{vec } \Delta_0 + \sum_{i=1}^N (x_i \otimes I_{m(k+g)}) \Sigma^{-1} \theta_i \right) + C \right) \right] \\ &= \exp \left[-\frac{1}{2} \left(\text{vec } \Delta' \left((\Omega^{-1} \otimes \Sigma^{-1}) + \left(\sum_{i=1}^N x_i x_i' \otimes \Sigma^{-1} \right) \right) \text{vec } \Delta \right. \right. \\ &\quad \left. \left. - 2 \text{vec } \Delta' \left((\Omega^{-1} \otimes \Sigma^{-1}) \text{vec } \Delta_0 + \sum_{i=1}^N (x_i \otimes \Sigma^{-1} \theta_i) \right) + C \right) \right] \quad (16) \end{aligned}$$

where C is a normalization constant that does not depend on Δ .

Equation (16) is proportional to a multivariate normal distribution, and it is of the same form that one sees in the corresponding step of a hierarchical multivariate regression (Rossi et al. 2005, sec. 2.12). The

conditional posterior covariance is easily identified as

$$\begin{aligned}\text{cov}(\text{vec } \Delta | \cdot) &= \left((\Omega^{-1} + X'X) \otimes \Sigma^{-1} \right)^{-1} \\ &= (\Omega^{-1} + X'X)^{-1} \otimes \Sigma\end{aligned}\tag{17}$$

The conditional posterior mean is also readily extracted from Equation (16), and after some tedious and mechanical manipulations can be simplified as

$$\begin{aligned}E(\text{vec } \Delta | \cdot) &= \left((\Omega^{-1} + X'X)^{-1} \otimes \Sigma \right) \left((\Omega^{-1} \otimes \Sigma^{-1}) \text{vec } \Delta_0 + \sum_{i=1}^N (x_i \otimes \Sigma^{-1} \theta_i) \right) \\ &= \text{vec} \left((\Delta_0 \Omega^{-1} + \Theta'X) (\Omega^{-1} + X'X)^{-1} \right)\end{aligned}\tag{18}$$

As Rossi et al. (2005) discuss, one might be tempted to simulate $\text{vec } \Delta$ directly from a multivariate normal distribution, with this mean and covariance. But repeated implementation of the Kronecker product is computationally inefficient, and the other steps in the algorithm require Δ , not $\text{vec } \Delta$ (we don't want to have to keep switching back and forth between these two forms of the same data). Fortunately, they propose a method to get to the *matrix* Δ in a smaller number of steps, without needing to do Kronecker multiplications.

Let Λ be the lower Cholesky root of Ω^{-1} , and let L be the lower Cholesky root of Σ^{-1} . Define W as a matrix that stacks X and Λ' , so $W = \begin{pmatrix} X \\ \Lambda' \end{pmatrix}$ is a $N + rJ \times p$ matrix. Then, $W'W = X'X + \Omega^{-1}$. Next, let $R_{W'W}$ be the Cholesky root of $W'W$. We can then write the conditional posterior covariance of $\text{vec } \Delta$ as

$$\begin{aligned}\text{cov}(\text{vec } \Delta | \cdot) &= (X'X + \Omega^{-1})^{-1} \otimes \Sigma \\ &= (W'W)^{-1} \otimes (LL')^{-1} \\ &= (R_{W'W} R_{W'W}')^{-1} \otimes (LL')^{-1} \\ &= (R_{W'W}^{-1} \otimes L^{-1})' (R_{W'W}^{-1} \otimes L^{-1})\end{aligned}\tag{19}$$

Next, let Ψ be an $rJ \times p$ matrix of independent standard normal draws. We can transform the random matrix Ψ into a posterior draw of $\text{vec } \Delta$ by multiplying Ψ by the Cholesky root of $\text{vec } \Delta$'s posterior covariance, and then adding its posterior mean. Some more mechanical manipulations get us to a posterior draw

of Δ .

$$\begin{aligned}
\text{vec } \Delta &= \text{vec} \left(\left(\Delta_0 \Omega^{-1} + \Theta' X \right) \left(\Omega^{-1} + X' X \right)^{-1} \right) + \left(R_{W'W}^{-1} \otimes L^{-1} \right)' \text{vec } \Psi \\
&= \text{vec} \left(\left(\Delta_0 \Omega^{-1} + \Theta' X \right) \left(\Omega^{-1} + X' X \right)^{-1} \right) + \text{vec} \left(L'^{-1} \Psi R_{W'W}^{-1} \right) \\
\Delta &= \left(\Delta_0 \Omega^{-1} + \Theta' X \right) \left(\Omega^{-1} + X' X \right)^{-1} + L'^{-1} \Psi R_{W'W}^{-1}
\end{aligned} \tag{20}$$

This approach avoids the need to compute Kronecker products and vec operators at each sweep of the Gibbs sampler. Also, $\Omega^{-1} + X'X$ is a constant (prior plus data), so it and $R_{W'W}$ need to be computed only once.

A.3 Sampling Σ

The conjugate prior for Σ is an inverse Wishart distribution, with ν degrees of freedom and location parameter A . The parameterization we use for the prior is

$$\pi(\Sigma | \nu, A) \propto |\Sigma|^{-\frac{\nu+m(k+g)+1}{2}} \exp \left[-\frac{1}{2} \text{tr} \left(A \Sigma^{-1} \right) \right] \tag{21}$$

The conditional posterior distribution for Σ is found multiplying Equations (13), (14), and (21). The result is an inverse Wishart distribution with $\nu + p + N$ degrees of freedom and a location parameter $A + \left[(\Delta - \Delta_0) \Omega^{-1} (\Delta - \Delta_0)' \right] + (\Theta - x\Delta')' (\Theta - x\Delta')$. A strategy for simulating from an inverse Wishart distribution is presented in Rossi et al. (2005, sec. 2.12).

A.4 Multiple Imputation of Missing Data

For customers who churn, but for whom there is no recorded reason for churn, we add an imputation step at the start of each Gibbs sweep. We treat the index of each missing cause of churn as an unknown parameter. By sampling from the conditional posterior distribution of this parameter at each Gibbs sweep, we essentially integrate over its marginal posterior distribution. Not only does this approach let us use the information we do have in those customer records, but we can also ensure that our estimates of the posterior distributions of the parameters of interest are not biased by the removal of incomplete records.

The conditional posterior of person i 's missing cause of churn is proportional to his data likelihood in Equation (3), times a prior on probabilities for the "true" reasons for churn. For simplicity, we use a multinomial prior with equal weights on all risks. The resulting conditional posterior is not of a standard form, but it is easily sampled from with a Metropolis-Hastings step. We used the prior distribution as

our proposal distribution as well. Note that if the proposal distribution does not place equal weight on all possible risks, the Metropolis-Hastings acceptance probability needs to be adjusted accordingly.

B Derivation of Equation (8)

Starting from Equation (7), the marginal effect of a single risk-specific parameter θ_j on ECLV is

$$\frac{\partial ECLV}{\partial \theta_j} = \sum_{t=1}^{\infty} \delta^t \frac{\partial S(t|\theta)}{\partial \theta_j} \quad (22)$$

Decompose $S(t)$ into its risk-specific components, and differentiate.

$$\begin{aligned} S(t|\theta) &= \exp \left[\log S_j(t|\theta_j) + \sum_{k \neq j} \log S_k(t|\theta_k) \right] \\ \frac{\partial S(t|\theta)}{\partial \theta_j} &= S(t|\theta) \frac{1}{S_j(t|\theta_j)} \frac{\partial S_j(t|\theta_j)}{\partial \theta_j} \end{aligned} \quad (23)$$

Then substitute (23) into (22).

C Adding μ_j to the remaining median lifetime

In this section we describe how to add μ_j to the median remaining risk-specific lifetime for risk j . For notational simplicity in this appendix, we suppress the j subscript, since we would only be working with one risk-specific distribution at a time. When the risk-specific timing distributions are median-parameterized Weibull distributions,

$$S(t|m, c) = 2^{-\left(\frac{t}{m}\right)^c} \quad (24)$$

where m is the median of the total lifetime for risk j (the time from when the customer is acquired to when he churns from risk j) and c is the Weibull shape parameter. For a customer who has already survived T periods, the probability of surviving to period t is

$$S(t|m, c, t > T) = 2^{-\left(\frac{t-T}{m}\right)^c} \quad (25)$$

To find the median remaining lifetime for a customer who survived T periods, m^* , solve the equation

$S(m^*|m, c, t > T) = \frac{1}{2}$ to get

$$m^* = (m^c + T^c)^{\frac{1}{c}} \quad (26)$$

Now, we want to extend this lifetime by μ , and compute the survival probabilities for all future periods. The survival probabilities are parameterized in terms of median *total* lifetime, so we need to find the m_μ that corresponds to a median remaining lifetime of $m^* + \mu$. Solving for m_μ in terms of m^* ,

$$\begin{aligned} m^* + \mu &= \left(m_\mu^c + T^c\right)^{\frac{1}{c}} \\ m_\mu &= \left[\left(m^* + \mu\right)^c - T^c\right]^{\frac{1}{c}} \end{aligned} \quad (27)$$

Substituting (26) into (27) yields

$$m_\mu = \left(\left[(m^c + T^c)^{\frac{1}{c}} + \mu\right]^c - T^c\right)^{\frac{1}{c}} \quad (28)$$

This μ -adjusted median of the total lifetime depends only on the customer's elapsed tenure T , the "un-adjusted" total median lifetime m , and the number of addition periods of time added to the remaining lifetime, μ .

References

- Berger, Paul D., Nada I. Nasr. 1998. Customer lifetime value: Marketing models and applications. *Journal of Interactive Marketing* **12**(1) 17–30.
- Beyersmann, Jan, Aurelien Latouche, Anika Buchholz, Martin Schumacher. 2009. Simulating competing risks data in survival analysis. *Statistics in Medicine* **28** 956–971.
- Blattberg, Robert C., Gary Getz, Jacquelyn S. Thomas. 2001. *Customer Equity*. Harvard Business School Press, Boston.
- Bogomolova, Svetlana, Jenni Romaniuk. 2009. Brand defection in a business-to-business financial service. *Journal of Business Research* **62** 291–296.
- Bolton, Ruth N. 1998. A dynamic model of the duration of the customer's relationship with a continuous service provider: The role of satisfaction. *Marketing Science* **17**(1) 45–65.
- Bolton, Ruth N., P. K. Kannan, Matthew Bramlett. 2000. Implications of loyalty program membership and service experiences for customer retention and value. *Journal of the Academy of Marketing Science* **28**(1) 95–108.
- Borle, Sharad, Siddharth S. Singh, Dipak C. Jain. 2008. Customer lifetime value measurement. *Management Science* **54**(1) 100–112.
- Chintagunta, Pradeep K. 1998. Inertia and variety seeking in a model of brand purchase timing. *Marketing Science* **17**(3) 253–270.

- Ehrenberg, Andrew S. C., Gerald J. Goodhardt, T. Patrick Barwise. 1990. Double jeopardy revisited. *Journal of Marketing* 52(2) 82–91.
- Ehrenberg, Andrew S. C., Mark D. Uncles, Gerald J. Goodhardt. 2004. Understanding brand performance measures: Using Dirichlet benchmarks. *Journal of Business Research* 57 1307–1325.
- Fader, Peter S., Bruce G. S. Hardie. 2009. Probability models for customer base analysis. *Journal of Interactive Marketing* 23 61–69.
- Fader, Peter S., Bruce G. S. Hardie. 2010. Customer-base valuation in a contractual setting: The perils of ignoring heterogeneity. *Marketing Science* 29(1) 85–93.
- Fader, Peter S., Bruce G. S. Hardie, Yuzhou Liu. 2009. Customer retention dynamics in a contractual setting: The paradox of increasing loyalty. Working paper, Wharton School, University of Pennsylvania.
- Gelman, Andrew, Xiao-Li Meng, Hal Stern. 1996. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica* 6(4) 733–807.
- Gupta, Sunil, Donald R. Lehmann, Jennifer Ames Stuart. 2004. Valuing customers. *Journal of Marketing Research* 41(1) 7–18.
- Gustafsson, Anders, Michael D. Johnson, Inger Roos. 2005. The effects of customer satisfaction, relationship commitment dimensions, and triggers on customer retention. *Journal of Marketing* 69 210–218.
- Han, Aaron, Jerry A. Hausman. 1990. Flexible parametric estimation of duration and competing risk models. *Journal of Applied Econometrics* 5(1) 1–28.
- Hoel, David G. 1972. A representation of mortality data by competing risks. *Biometrics* 28(2) 475–488.
- Hogan, John E., Katherine N. Lemon, Roland T. Rust. 2002. Customer equity management: Charting new directions for the future of marketing. *Journal of Service Research* 5(1) 4–12.
- Jones, Thomas O., W. Earl Sasser. 1995. Why satisfied customers defect. *Harvard Business Review* 73 88–99.
- Keaveney, Susan M. 1995. Customer switching behavior in service industries: An exploratory study. *Journal of Marketing* 59(2) 71–82.
- Keiningham, Timothy, Terry G. Vavra, Lerzan Aksoy, Henri Wallard. 2005. *Loyalty Myths*. John Wiley and Sons, Inc., Hoboken, N.J.
- Lehmann, Donald R. 2004. Metrics for making marketing matter. *Journal of Marketing* 68 73–75.
- Little, Roderick J. A., Donald B. Rubin. 2002. *Statistical Analysis with Missing Data*. 2nd ed. Wiley-Interscience, Hoboken, N.J.
- Moe, Wendy W., Michael Trusov. forthcoming. Measuring the value of social dynamics in online product ratings forums. *Marketing Science* .
- Oliver, Richard L. 1999. Whence customer loyalty? *Journal of Marketing* 63 33–44.
- O’Sullivan, Don, Andrew V. Abela. 2007. Marketing performance measurement ability and firm performance. *Journal of Marketing* 71 79–93.
- Prentice, R. L., J. D. Kalbfleisch, A. V. Peterson, N. Flournoy, V. T. Farewell. 1978. The analysis of failure times in the presence of competing risks. *Biometrics* 34(4) 541–554.
- Putter, H., M. Fiocco, R. B. Geskus. 2007. Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine* 26 2389–2430.
- Raghubir, Priya, Sanjiv R. Das. 2010. The long and short of it: Why are stocks with shorter runs preferred? *Journal of Consumer Research* 36 964–982.

- Reichheld, Frederick F., W. Earl Sasser. 1990. Zero defections: Quality comes to services. *Harvard Business Review* **68** 105–111.
- Reinartz, Werner, Jacquelyn S. Thomas, V. Kumar. 2005. Balancing acquisition and retention resources to maximize customer profitability. *Journal of Marketing* **69** 63–79.
- Reinartz, Werner J., V. Kumar. 2003. The impact of customer relationship characteristics on profitable lifetime duration. *Journal of Marketing* **67**(1) 77–99.
- Rossi, Peter E., Greg M. Allenby, Robert McCulloch. 2005. *Bayesian Statistics and Marketing*. Wiley Series in Probability and Statistics, John Wiley and Sons, Chichester, UK.
- Rubin, Donald B. 1976. Inference and missing data. *Biometrika* **63**(3) 581–592.
- Rubin, Donald B. 1984. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Annals of Statistics* **12**(4) 1151–1172.
- Rust, Roland T., Katherine N. Lemon, Valarie A. Zeithaml. 2004. Return on marketing: Using customer equity to focus marketing strategy. *Journal of Marketing* **68**(1) 109–127.
- Rust, Roland T., Anthony J. Zahorik, Timothy Keiningham. 1995. Return on quality (ROQ): Making service quality financially accountable. *Journal of Marketing* **59**(2) 58–70.
- Schweidel, David A., Peter S. Fader, Eric T. Bradlow. 2008a. A bivariate timing model of customer acquisition and retention. *Marketing Science* **27**(5) 829–843.
- Schweidel, David A., Peter S. Fader, Eric T. Bradlow. 2008b. Understanding service retention within and across cohorts using limited information. *Journal of Marketing* **72** 82–94.
- Seetharaman, P. B., Pradeep K. Chintagunta. 2003. The proportional hazard model for purchase timing: A comparison of alternative specifications. *Journal of Business and Economic Statistics* **21**(3) 368–382.
- Sharp, Byron. 2010. *How Brands Grow*. Oxford University Press, Melbourne.
- Srinivasan, Raji, Gary L. Lilien, Arvind Rangaswamy. 2008. Survival of high tech firms: The effects of diversity of product-market portfolios, patents and trademarks. *International Journal of Research in Marketing* **25** 119–128.
- Venkatesan, Rajkumar, V. Kumar. 2004. A customer lifetime value framework for customer selection and resource allocation strategy. *Journal of Marketing* **68**(4) 106–125.
- Verhoef, Peter C. 2003. Understanding the effect of customer relationship management efforts on customer retention and customer share development. *Journal of Marketing* **67**(4) 30–45.
- Vilcassim, Naufel J., Dipak C. Jain. 1991. Modeling purchase-timing and brand-switching behavior incorporating explanatory variables and unobserved heterogeneity. *Journal of Marketing Research* **28**(1) 29–41.