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

May 10, 2011

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 this framework, we demonstrate that the impact of a firm's efforts to reduce customer churn for controllable reasons is mitigated by the prevalence of uncontrollable ones, resulting in a "damper effect" on the return from a firm's retention marketing efforts. We use data from a provider of land-based telecommunication services to demonstrate how the competing risk model can be used to derive a measure of the incremental customer value that a firm can expect to accrue through its efforts to delay churn, taking this damper effect into account. In addition to varying across customers based on geodemographic information, the magnitude of the damper effect depends on a customer's tenure to date. 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 retention efforts should be deployed.

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 is irrelevant. All that matters is that, when a customer churns, the revenue stream stops.

A separate stream of research has examined how factors such as customers' satisfaction with service and commitment relate to customer retention (Bolton 1998; Verhoef 2003; Gustafsson et al. 2005). Keaveney (1995) investigates how certain critical incidents caused customers to switch from one service provider to another. Surveying more than 500 service customers, she identifies eight categories of critical incidents for which customers switch service providers: pricing, inconvenience, core service failures, service encounter failures, response to service failure, competition, ethical problems, and involuntary switching. Though a number of these reasons potentially can be influenced by the firm, such as those relating to service failures, pricing or competition, others such as involuntary switching are outside of the firm's control. More recently, Bogomolova and Romaniuk (2009) surveyed 765 business-owners who canceled their electronic funds transfer services and found that 57% of the cancellations were for reasons beyond the bank's control. Although 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 service, or even death. We define those reasons for which customers may churn that cannot reasonably be influenced by the firm's efforts directly as "uncontrollable." Sharp (2010) describes how uncontrollable churn can lead to misleading brand loyalty metric, since the defection of some customers may have nothing to do with the quality of the product or service.

Though prior research has examined the reasons for which customers churn, this information has yet to be incorporated into quantitative models of customer lifetime or residual customer valuation. This raises a key question: although there are multiple reasons for which customers may end their relationship with a service provider, does it 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 discounted expected future revenue streams (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 need to know why customers churn so that they may pull the appropriate marketing or operational “levers” to delay churn. 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). If customers are switching to service providers who are seen to offer a more appealing service for a better price, the firm might instead engage in marketing activities that emphasize its own perceived value proposition.

Although service quality and perceived value can be influenced by the firm, what about those reasons for churn that are beyond the control of the firm, such as death, relocation, or unexpected changes in economic circumstances? It is unlikely that the firm’s marketing efforts could delay such events or the ensuing churn. The challenge for managers is to identify those individuals who are most prone to churn for reasons under the firm’s influence. Doing so would allow firms to focus their efforts on customers for whom their actions have the greatest impact. Incorporating the reasons for churn into a 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 also enable the firm to determine how much return it can expect to recoup from its customer retention efforts through an examination of the expected change in customers’ *remaining* lifetimes. Such information offers guidance in terms of when to expend effort to increase retention and to which of the firm’s existing customers (Venkatesan and Kumar 2004), thereby contributing to the development of a dynamic strategy (Hogan et al. 2002; Rust et al. 2004; Reinartz et al. 2005).

In this research, we contribute to the customer retention and lifetime value literature by examining how the likelihoods of churning for different reasons (i.e., controllable or uncontrollable) vary over time and across customers. By decomposing the likelihood of customer churn into the different reasons for which it occurs, firms can adopt the appropriate strategies to increase customer retention. These tactics will differ not only across customers, but also over the course of customers’ relationships, depending on the relative prevalence of different reasons for churn. To quantify this, we extend the idea of discounted expected residual lifetime (DERL, Fader and Hardie (2010)) by deriving the *incremental* DERL as a measure of return on investment in retention marketing. We show that while a firm may take steps to reduce churn due to reasons under its control, the presence of uncontrollable reasons for churn creates a *damper effect* on the potential impact of the firm’s actions. Depending on the prevalence of uncontrollable churn, the expected increase in the length of a customer’s relationship (and hence the increase in the revenue the firm potentially can generate) may be substantially diminished. The strength of this damper effect need not be constant, but rather may change over time as the likelihoods of churning due to controllable and uncontrollable reasons shift with a customer's tenure. Since customers' latent propensities to churn are unobserved, we employ Bayesian updating to infer posterior estimates of incremental DERL that depend only on observable data: the elapsed tenure of the relationship.

Our examination of DERL and the damper effect is consistent with the spirit of the economic research on behavior-based price discrimination that proposes charging customers' different prices based on their past behavior that is observable to the firm (Shaffer and Zhang 2000; Taylor 2003). Though this research has focused on pricing, Fudenberg and Villas-Boas (2007) note the potential for firms to leverage information on customers' past behaviors for other decisions such as targeted communications and promotions, which can be used to reduce churn (Shaffer and Zhang 2002). In line with this work, we take advantage of the available information from customers' past behaviors (i.e., the decision to maintain service) to update our beliefs of when customers will churn *and* whether this will be due to a controllable or uncontrollable reason. If those reasons for churn that the firm can influence become more likely the longer that a customer has maintained service, the firm may see larger returns from retention marketing by taking actions later in customer relationships.

The remainder of this article proceeds as follows. In Section 2, we describe a competing risk model to jointly model the duration for which a customer maintains a contractual service and the reason for which he cancels it, and then demonstrate the damper effect due to the presence of multiple causes of churn. We then introduce the idea of incremental DERL as a measure of return on investment in retention activities. This framework is appropriate for any contractual customer relationship setting, such as subscription-based services (e.g., cable television), ongoing memberships (e.g., health club or food co-op memberships), or automatic direct debit charitable donations.[1] In Section 3, we offer an example of the competing risk framework in action, using data from a provider of land-based telecommunications services. In this example, we show how combining the ideas of competing risks, incremental DERL, and Bayesian updating, can yield novel insights for researchers and practitioners. We conclude with a discussion of the limitations of the current work and directions for future research in Section 4.

[1] By "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); the key distinction is that the customer must act directly to terminate the relationship.

2 A Competing Risk Model of Customer Retention

We use a hierarchical competing risk framework to model jointly the duration of the customer relationship with a firm and the reason for which the relationship is terminated. 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 (the reason that the customer says triggered the event). We refer to these causes as "risks" because a customer is "at risk" of churning from one of them. Competing risk models are common in medical fields such as medicine and epidemiology (Putter et al. 2007), where there are multiple potential causes of death but only one may be observed. In marketing, Vilcassim and Jain (1991) and Chintagunta (1998) employ a competing risk setup to jointly model inter-purchase time and brand switching behavior. Srinivasan et al. (2008) use a competing risk model to examine the survival of high tech firms, and Moe and Trusov (2010) 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.

Extant models of contractual customer retention 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 time (Schweidel et al. 2008a; Fader and Hardie 2010; Borle et al. 2008). Extending customer retention models to a competing risk framework, we let the time-to-churn from controllable and uncontrollable risks be governed by their own processes, and we assume that these processes are *conditionally* independent. Placing a hierarchical structure on the basic competing risk framework lets us incorporate both observed and unobserved heterogeneity, and, as in Moe and Trusov (2010), to allow for correlations across customers in the risk-specific churn propensities (e.g., customers who are more susceptible to churning from a controllable risk may also be more susceptible to churning from an uncontrollable risk). This is consistent with incorporating an unobserved covariate to correlate the propensities to churn for different reasons, as discussed by Clayton and Cuzick (1985).[2]

[2]While we employ correlated risk-specific parameters, another way in which a correlation can be induced is by using a copula with a nonparametric baseline hazard function, and with observed regressors, as discussed by Heckman and Honoré (1989) and Abbring and Berg (2003). We favor the use of a parametric competing risk model with correlated parameters, as the additional structure allows us to generate forecasts beyond the calibration period.

2.1 Model Specification

Our hierarchical competing risk model has three levels: the data likelihood, conditional on individual- and risk-specific parameters; the prior specification on those parameters; and the hyperpriors on the prior parameters. For the data likelihood, it is instructive to start with the special case of a single risk, which is a standard censored timing model. The observed outcome variable for person i (out of a population of size N) is the vector $[t_i, d_i]$, where t_i is the duration for which i is observed to retain service during the calibration period and d_i indicates an uncensored observation. If i drops service during the calibration period, then $d_i = 1$. Should i maintain service throughout the calibration period, t_i is equal to the duration for which i is under observation (T_i) and $d_i = 0$. The survival probability $S(t_i|\theta_i)$ is the probability that i churns sometime after time t_i, conditional on a parameter vector θ_i. The choice of $S(t_i|\theta_i)$ leads to functions of θ_i that represent our beliefs about values of interest related to the individual's lifetime. For example, if $S(t_i|\theta_i) = e^{-\theta_i t_i}$ (an exponential distribution), then $1/\theta_i$ is that person's expected time-to-churn. The survival probability is related to the cumulative distribution function by $F(t_i|\theta_i) = 1 - S(t_i|\theta_i)$.

Throughout the paper we maintain an assumption that churn occurs in continuous time, but is observed in discrete time. Therefore, the data likelihood for a single-risk timing model is, for right-censored observations, the probability of the individual surviving to the end of the observation period. For uncensored observations, the data likelihood is the probability of surviving to time $t_i - 1$, but having churned before time t_i. Put together, the single-risk data likelihood is

$$L(t_i, d_i|\theta_i, T_i) = \left[S\left(t_i - 1|\theta_i\right) - S\left(t_i|\theta_i\right)\right]^{d_i} S(T_i|\theta_i)^{1-d_i} \tag{1}$$

Competing risk models generalize such models by allowing for multiple causes of churn. The intuition is that churn could potentially occur due to one of several different risks, with the time of churn due to a particular risk being governed by a risk-specific random process. For competing risk models, as in the single-risk model, the observed data includes a churn/censoring time t_i, as well as the non-censoring indicator d_i. Additionally, for uncensored observations, we observe j_i, which is an index to the one of the J possible risks that caused the churning event. We then define $F_j(t_{ij}|\theta_{ij})$ as the probability of churning *from risk* j before time t_{ij}. Although we model J distinct t_{ij}'s, we observe only one of them; $t_i = t_{ij}$ for the j that indexes the risk that was observed to have occured, while t_{ij} for any other j is right-censored. Letting $S_j(t_{ij}|\theta_{ij}) = 1 - F_j(t_{ij}|\theta_{ij})$ denote the corresponding risk-specific survival probabilities, the *aggregate* survival probability $S(t_i|\theta_i)$ is the probability of surviving *all* possible risks through time t_i (θ_i is the concatenation of all J of the risk-specific θ_{ij} vectors). Thus, the data likelihood under a competing risk framework is, for uncensored observations, the probability of both surviving all risks through time $t_i - 1$, and then churning from risk j_i at some time between times $t_i - 1$ and t_i. For censored observations, the data likelihood is the probability of surviving all risks through the censoring time. The data likelihood for the competing risk model can be written as

$$L(t_i, j_i, d_i|\theta_i, T_i) = \left[\left(S_j\left(t_i - 1|\theta_{ij}\right) - S_j\left(t_i|\theta_{ij}\right)\right) S(t_i - 1|\theta_i)\right]^{d_i} S(T_i|\theta_i)^{1-d_i} \tag{2}$$

The key notational difference is that under a competing risk model, there are, for each person, J risk-specific survival probabilities, $S_j(t_i|\theta_{ij})$, each controlled by its own person- and risk-specific parameter θ_{ij}. The probabilities of churning for each of the risks are assumed to be independent, conditional on θ_i. Correlation in these propensities to churn for each risk is induced through correlations in the parameters, which we discuss below.

The functional forms for each $S_j(\cdot)$ are prespecified as part of the model. The data likelihood in Equation 2 depends not only on the risk-specific survival probabilities $S_j(t)$, but also the *aggregate* survival probabilities $S(t)$. To get $S(t)$, we use the risk-specific hazard rate functions $H_j(t|\theta_{ij})$, which capture the probability of churning at time t conditional on not having churned by time $t-1$. In the case of controllable churn, the associated hazard rate function may be related to a customer's level of satisfaction and prior experience (Bolton 1998), which may change over the duration of the relationship. An increasing hazard rate may result, for example, from an accumulation of negative experiences with a service provider. As not all customers will have the same set of experiences or experience them at the same rates, we allow for heterogeneity in the hazard rates across customers.[3]

It is 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):

$$S(t|\cdot) = \exp\left[-\sum_{t'=1}^{t} H(t'|\cdot)\right] \tag{3}$$

When there is more than one risk, the aggregate hazard rate $H(t|\theta_i)$ at time t is simply the sum of the risk-specific hazard rates (Prentice et al. 1978).

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

A generally applicable way to derive the aggregate survival probability is to substitute Equation 4 into

[3] As we discuss in Section 4, data on customers' experiences can be incorporated into our modeling framework when such data are available.

Equation 3. This method is particularly useful in the presence of time-varying covariates, such as with proportional hazard regression. Alternatively, when the hazard rate depends only on the elapsed time of the processes, we can derive the aggregate survival probability directly from the joint probability of the risk-specific survival functions.

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

The next stage of the hierarchical model is the prior distribution $g(\theta_i)$. Note that each element of θ_i is one parameter for one of the risk-specific survival probabilities. Some customers may be more likely to churn than other customers, and some customers may be more prone to certain risks than to other risks. Additionally, these differences may be due to both observable factors, such as demographics, and unobservable ones. A convenient way to express this heterogeneity in θ_i is through a multivariate generalized linear model in which Δ is a matrix of coefficients that maps an observable factor vector x_i to θ_i, and Σ is a covariance matrix.

$$g\left(\theta_i\right) \sim MVN\left(\Delta x_i, \Sigma\right) \tag{6}$$

Correlation in risk-specific churn processes to is captured through the off-diagonal elements of posterior estimate of Σ. This correlation is across individuals, allowing those who are more prone to one risk to also be more (or less) prone to another risk. For example, we might find that customers who are more likely to move out of the service area are also the customers who are most price sensitive.

The third level of the hierarchical model consists of the hyperprior π on Δ and Σ. Since we we have little prior knowledge about these parameters, our choices arise primarily out of a desire for computational convenience. Our prior on Δ is placed on $\operatorname{vec}\Delta$ (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 $\operatorname{vec}\Delta$), 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 and ν degrees of freedom, scaled such that the *priors* on the elements of θ_i are uncorrelated and weakly informative. However, note that this model does allow for posterior correlation among the parameters of the J risk-specific models. Therefore, we take into account the possibility that some individuals who are more prone to one risk may be more, or less, prone to another risk.

All together, the posterior distribution of unknown parameters is

$$g(\theta_{1:H}, \Delta, \Sigma | x_{1:N}, A, \nu, \Delta_0, \Omega) \propto \prod_{i=1}^{N} \Big[L(t_i, d_i, j_i | \theta_i)\; g(\theta_i | x_i, \Delta, \Sigma) \Big] \times \pi(\Delta, \Sigma | \Delta_0, \Omega, \nu, A) \tag{7}$$

where $L(\cdot)$ is defined in Equation 2.

2.2 The Damper Effect on Customer Retention and Lifetime Value

Conditional on churning at time t_i, the relative risk of churning due to risk j is the ratio of the hazard rates: $H_j(t|\theta_{ij}) / \sum_{j'=1}^{J} H_{j'}(t|\theta_{ij'})$(Beyersmann et al. 2009). When the risk-specific hazards change over time (i.e., in the presence of risk-specific duration dependence), the probability with which a customer churns due to a particular risk will also evolve. Such changes in the relative risks are 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 a veteran customer is more likely to churn due to a reason the firm can influence compared to a newer customer, the firm may find it is better served by shifting its retention resources toward older customers.

To highlight these practical implications, next we discuss the metrics that can be derived from our hierarchical competing risk model for use in managing customer retention. In this section, we demonstrate how changes in individual-level risk-specific propensities for churn (characterized by the elements of θ_i) can affect expected customer lifetime value (ECLV). We define ECLV as the stream of future expected cash flows, discounted back to the present, that the firm can expect to accrue from a newly acquired customer. Assuming a \$1-per-time-period future revenue stream from a single customer, and a known, constant and homogeneous discount factor δ, ECLV is:

$$ECLV\,(\theta_i) = \sum_{t=1}^{\infty} S\,(t|\theta_i)\,\delta^t \tag{8}$$

If a firm were to "slow down" the churn rate of its customers, this will lead to an increase in future survival probabilities and consequently customer tenure, ultimately driving an increase in ECLV.

But what would happen if we were to delay churn that is attributable to a specific risk the customer faces? In Appendix B, we show that the impact of a change in θ_{ij} on ECLV is:

$$\frac{\partial ECLV(\theta_i)}{\partial \theta_{ij}} = \sum_{t=1}^{\infty} \delta^t \frac{S(t|\theta_i)}{S_j(t|\theta_{ij})} \frac{\partial S_j(t|\theta_{ij})}{\partial \theta_{ij}} \tag{9}$$

The factor on the right in the summand in Equation 9 is how much the risk-specific survival probability at time t changes per unit change in θ_{ij}, holding all of the other elements of θ_i constant (i.e., the marginal effect of θ_{ij} on $S_j(t)$). As an example, if θ_{ij} were the median time-to-churn from risk j, then this marginal effect factor would be positive. The middle factor is the ratio between the aggregate and risk-specific survival probabilities. If we are considering only a single risk, then by definition $S_j(t|\theta_{ij}) = S(t|\theta_i)$ and this ratio is 1. Therefore, the full impact of the marginal effect manifests as an improvement in ECLV. When a customer faces more than one risk, however, the ratio of $S(t|\theta_{ij})$ to $S_j(t|\theta_i)$ is less than 1. This ratio acts as a *damper* on efforts to slow down churn due to the j^{th} risk. If risk j' is an unlikely cause of churn in the population, then the damper ratio of $S(t|\theta_{ij})$ to $S_j(t|\theta_i)$ will be close to 1, and 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 more likely to occur, this ratio will be lower and consequently reducing churn due to risk j will have a diminished impact on ECLV. In the presence of duration dependence (i.e., non-constant hazard rate functions) in the risk-specific hazard functions, the probability of a customer churning from different risks could change as the duration of the relationship with the company increases, which can result in changes in the damper ratio.

The magnitude of the damper effect does not, by itself, tell a manager whether or not he should invest in retention marketing. If the cost of investment is low, and the per-period revenue is high, then the net present value of the investment may still be positive, even if the incidence of the controllable churn is rare. When faced with the decision to make a substantial investment in reducing churn due to risk j, however, the impact of this damper effect may suggest that such an investment in retention marketing is not warranted due to the high prevalence of the other reasons for churn.

A customer's expected tenure is given by the infinite sum of the survival function, and so the expected lifetime can be thought of as the area underneath the survival curve. To illustrate the damper effect, consider a service for which there are only two risks, one controllable and one uncontrollable, and let the risk-specific survival functions exhibit positive duration dependence. In Figure 1, we illustrate two possible survival curves. The solid line is the aggregate survival curve in the presence of both controllable and uncontrollable risks, and the area under that curve represents the expected lifetime of that customer. The dashed line represents the survival curve if the firm could completely eliminate churn due to the controllable risk. In the presence of uncontrollable reasons for churn, even if the firm were to eliminate all controllable churn, the resulting survival function becomes the risk-specific survival function associated with uncontrollable churn. Thus, the area between the two curves is the incremental lifetime that the firm could accrue if it were to eliminate all controllable churn.

An alternative way of illustrating the effect of reducing controllable churn is to consider how the

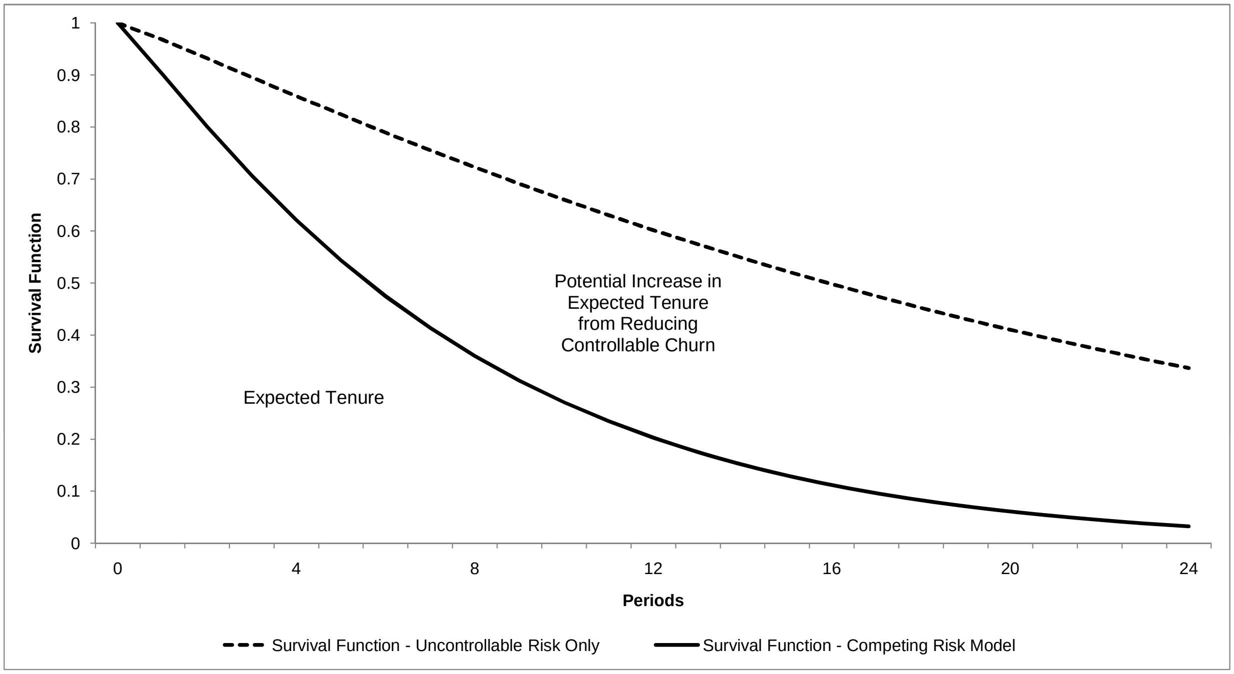

Figure 1: Maximum expected tenure in the presence of uncontrollable churn

hazard and survival functions change if the firm were able to delay the time until customers churn. To operationalize this, we examine the impact of increasing the median time-to-churn for a single customer by μ periods. We start with a case in which there is a 1:2 ratio between the median time-to-churn from controllable and uncontrollable risks. In Figure 2, we illustrate the impact of extending the median time-to-churn for the controllable risk by 4, 8, and 12 time periods. The left panels depict the corresponding changes in the hazard function. As we increase the median time-to-churn, the prevalence of the controllable risk falls, shifting its hazard function downward, while the hazard function for the uncontrollable risk is unchanged. The shaded area in the right panels show the incremental change in customers' expected tenures resulting from increasing the median time-to-churn. In the top-right panel of Figure 2, the black band is the increase in expected tenure resulting from increasing the median time-to-churn by 4 time periods. The middle-right shows the increase in expected tenure resulting from the initial increase of 4 time periods, as well as the incremental increase from extending the median time-to-churn by another 4 periods (for a total of 8 periods). In the bottom-right panel, we see the incremental increases for each of the three 4-period intervals.

Although the prevalence of controllable churn declines as we increase the median time-to-churn from the controllable risk by up to 12 periods, the presence of uncontrollable churn dampens the incremental improvements in expected tenure. In Figure 3, we depict the incremental increases in ECLV (Equation 8)

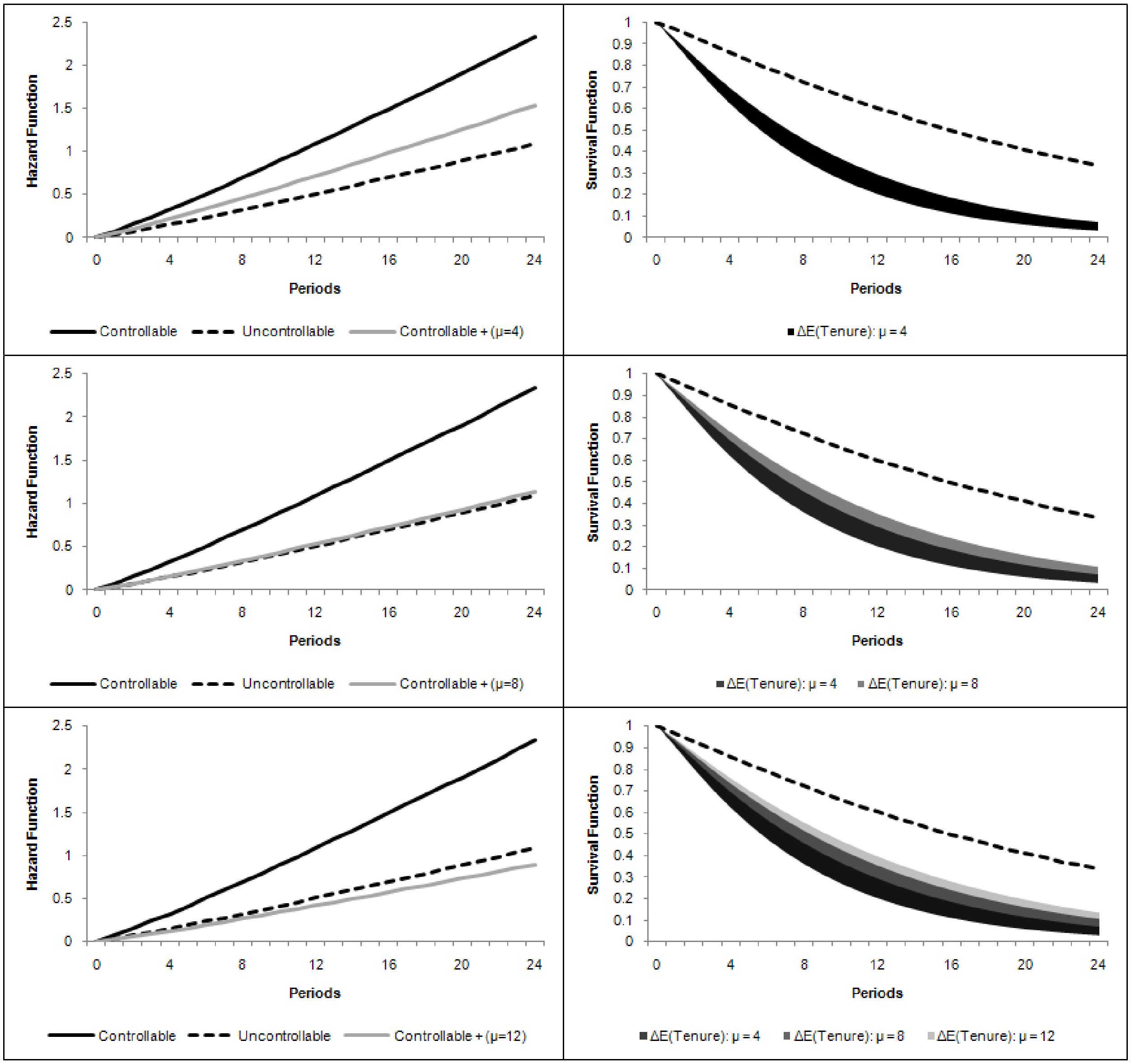

Figure 2: Impact of reducing controllable churn on expected tenure. As we delay controllable churn (holding the rate of uncontrollable churn constant), the hazard rate function falls, and the expected customer lifetime increases.

that the firm gets by extending the median time-to-churn for the controllable risk one period at a time. The highest line shows what the incremental increases would be in the absence of uncontrollable churn. The subsequent lines illustrate the additional incremental expected lifetime in the presence of uncontrollable churn. Each line corresponds to a different ratio of the median time-to-uncontrollable-churn to the median time-to-controllable-churn. These ratios are indicative of the relative prevalence of the risks before any intervention takes place, with larger ratios indicating scenarios in which uncontrollable churn is more likely. The presence of uncontrollable churn places an upper bound on the benefit (in terms of ECLV) of delaying controllable churn. The incremental increases in ECLV from curbing controllable churn are smaller when uncontrollable churn is more prevalent, even though the prevalence of controllable churn remains the same.

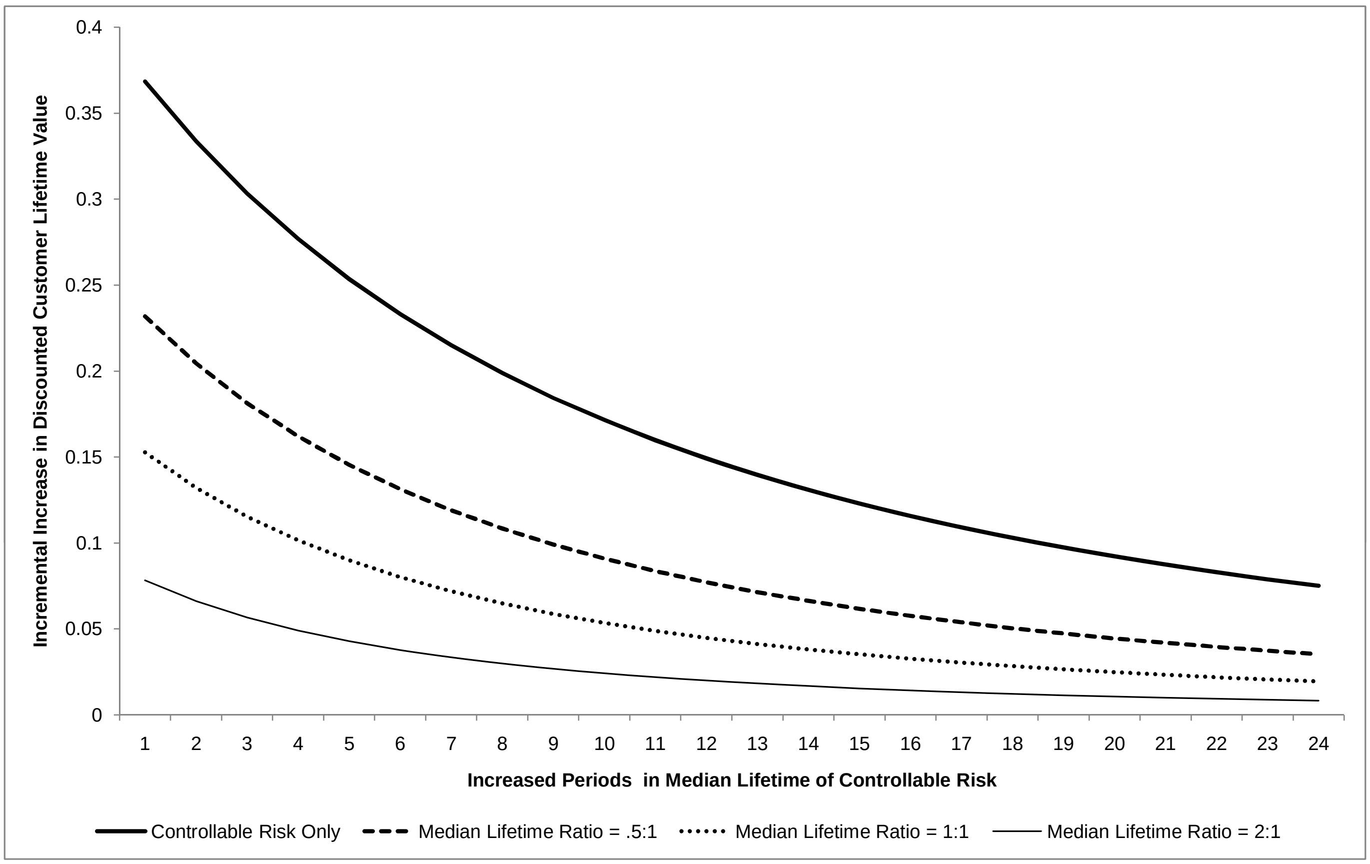

Figure 3: Damper effect varying the median time-to-churn due to the uncontrollable risk. The median lifetime ratios reflect the ratio of the median time-to-churn from the uncontrollable risk to the median time-to-churn from the controllable risk. The ratio of .5:1 corresponds to the scenario illustrated in Figures 1 and 2. Larger ratios indicate an increased likelihood of churning due to the uncontrollable risk, thereby dampening the incremental increases in customer lifetime associated with delaying controllable churn.

In this example, and in the empirical analysis that follows in Section 3, we focus on the distinction between controllable and uncontrollable churn. However, the damper effect will exist even if we have more than two risks, and when we use different definitions of the risks. Consider, for example, three of the categories of critical incidents identified by Keaveney (1995) that can potentially be influenced by the firm: core service failures, service encounter failures and response to service failures. Whether we treat these as three distinct risks or a single risk, the net effect of reducing service-related failures on overall churn will still depend on the prevalence of churn due to uncontrollable reasons (i.e., involuntary switching). Let us also consider switching due to inconvenience, another category identified by Keaveney (1995), which could potentially be influenced by the firm. For customers among whom churn due to inconvenience occurs frequently, the impact of reducing churn due to service-related failures may be limited, as they are prone to churn for a reason unrelated to service-related failures. Thus, the damper effect plays a role when considering two distinct sources of controllable churn. Each reason for churn j has a damper factor associated with it, $S(t)/S_j(t)$, that depends on the likelihoods of other reasons for churn. It is the presence of additional causes of churn–regardless of what these risks are–that creates the damper effect which limits the potential value of retention marketing.

2.3 Incremental DERL

In our formulation of the hierarchical competing risk model, individual-specific risk propensities θ_i depend on observed covariates and population-level parameters. Once we estimate Δ and Σ from historical data, we can estimate the value of a new customer by drawing a new θ_i from its prior in Equation 6. Indeed, to get predictive distributions of ECLV and the damper effect for a new customer, we integrate Equations 8 and 9 over the prior. When we do this, we assume that all of the unobserved variation in θ_i is driven by a purely random effect, and then once observed characteristics in x_i are taken into account, any brand new customer is a priori identical to another.

However, data from existing customers reveal an additional piece of information: T_i, the elapsed tenure of the customer relationship. Consider two customers with identical demographic profiles (i.e., they have the same x_i). One customer has been with the firm a long time (much longer than the average customer), and the other was acquired recently. The expected value of θ_i for each of these customers is $\Delta' x_i$. A customer who has been around a long time, however, probably has a lower propensity to churn than a newer customer. The additional information that we get by observing T_i for a customer lets us update our beliefs about how likely that customer is to churn, and consider posterior distributions of θ_i.

Fader and Hardie (2010) define DERL (discounted expected residual lifetime) as the future expected discounted cash flows that the firm accrues from a customer who has a relationship with the firm that is T_i periods old, and who will pays \$1 per period in each future period that the relationship is still active. The concept of DERL allows for a Bayesian updating strategy to exploit any additional information that is contained in T_i. In the presence of duration dependence, knowing T_i provides information on how much longer it might be until the customer decides to churn. Conditional on θ_i, DERL is:

$$DERL(T_i|\theta_i) = \sum_{t=T_i}^{\infty} \delta^{t-T_i+1} S(t|\theta_i, t > T_i) \tag{10}$$

The distinctionbetween Equations 8 and 10 is that Equation 10 conditions on the customer having already survived T_i periods. This means that customers with long elapsed tenure may be more (positive duration dependence) or less (negative duration dependence) likely to churn from a particular risk than customers who were acquired more recently. Whether duration dependence is positive or negative depends on the choices for S_j and the estimates of θ_i. To take into account the information we learn from T_i about θ_i, we integrate Equation 10 over i's posterior distribution $\theta_i|T_i, x_i, \Delta, \Sigma$, as opposed to integrating Equation 8 over the prior $\theta_i|x_i, \Delta, \Sigma$. This gives us an expression for DERL that depends only on population-level parameters (Δ and Σ) and observed individual data (T_i, x_i), but not the latent, unobserved variables θ_i.

$$DERL(T_i) = \int \sum_{t=T_i}^{\infty} \delta^{t-T_i+1} S(t|\theta_i, t > T_i)\; g(\theta_i|\Delta, \Sigma, x_i, t > T_i)\; d\Delta\; d\Sigma \tag{11}$$

Fader and Hardie (2010) derive a closed-form expression for DERL for a single-risk case, under different distributional assumptions and without demographic information. They also report a closed-form expression for "elasticity of retention," which is the percentage change in DERL that the firm can expect from a one percent increase in the retention rate. In the competing risk case, we can estimate Equation 11 numerically, and then compute an *incremental* DERL that results from a firm's retention marketing activities. We define incremental DERL as the difference between the DERL that we expect to accrue with no intervention, and the DERL that we would get if we could delay churn from risk j by μ_j periods. Incremental DERL, multiplied by the per-period revenue, yields the additional discounted cash flow that a firm could expect by delaying churn from risk j. Expenditures in retention marketing that are aimed at reducing churn associated with risk j can then be weighed against this measure of the expected return from the firm's investment. We will show in Section 3 that incremental DERL from delaying churn from risk j depends on the propensities for *all* reasons for churn, which are informed by a customer's time-invariant geodemographic cluster and his tenure to date. As we will demonstrate, it may require a substantial investment in reducing risk-specific churn to yield a small increase in a customer's value, due to the presence of the damper effect.

3 Empirical Example: A Telecommunications Service Provider

To demonstrate the use of the competing risk model, and to examine the damper effect and incremental DERL in one particular context, we now introduce an empirical example using data from a US-based telecommunications provider. The firm, which chose to remain anonymous, offers wired, subscription-based telecommunication services that are geographically based (service is brought into a home directly), and has been doing so for many years. 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 became customers between January, 2007 and March, 2008. Our observation period for these homes continues through June, 2008. For each household, we observe the month in which the customer was acquired and, if the customer churned while under observation, the month in which churn was observed. There were no substantive changes in the firm's or competitors' offerings during the observation period. It should be noted that our data, and hence our analysis, are at the customer level rather than the service level.[4].

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 of these categorical variables, we were left with 15 geodemographic variables for each household. As is commonly the case when segmenting customers based on their geographic location, we do not know with certainty if a particular household matches the demographic vector assigned to it based on its geographic location. Inferences about the marginal effect of any geodemographic variable should bear this in mind. As such data are commonly available and employed by service providers, we employ it in our analysis to highlight the managerial relevance of this research.

For 77% of those customers churn during the observation period, we also observe a customer-stated "reason for churn."[5] At the time of churn, customers reported one of eight different reasons for churn: Competition, Price, Dissatisfaction with Service, Dissatisfaction with Product, Moving from Service Area,

[4] In section 4, we discuss possible directions for future research if service level data were available

[5] We always assume that customers are telling the truth.

Death/Divorce/Family Issues, Nonpayment, and Abuse. The decision to switch to a competing service provider may stem from a source of dissatisfaction with the current provider, whether it is related to service price or quality. As such reasons for churn can reasonably be affected by the firm's activities, such as targetted promotions, we consider the decision to switch to a competitor to be a form of controllable churn. Consistent with the distinction between reactional triggers that affect perceptions of service performance (e.g., service failures) and situational triggers that are related to changes in customers' lives (e.g., personal or economic circumstances) made by Gustafsson et al. (2005), we consider the next three reasons for churn (price, dissatisfaction with service, dissatisfaction with product) as controllable and the following two reasons (moving from service area, death/divorce/family issues) as uncontrollable. Uncontrollable churn is also analagous to "involuntary switching" (Keaveney 1995). We placed the Nonpayment and Abuse reasons into a third category, called Nonpay-Abuse. Unlike the first two categories of churn reasons, churn due to nonpayment or abuse is initiated by the company rather than the customer. As firm-initiated churn likely arises from an underlying process that is distinct from the processes underlying customer-initiated churn, we consider it as a third type of risk.

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		
Controllable	1022	132
Uncontrollable	2239	259
Nonpay-Abuse	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.

We considered a more granular categorization scheme, such as separating the different descriptive reasons for controllable churn into separate categories. Indeed, the competing risk framework can support any number of risk categories. However, a customer's decision to switch to a competitor and his dissatisfaction with the price or service are not necessarily mutually exclusive. Negative experiences with the service provider may accumulate over time, resulting in lower satisfaction or calculative commitment, both which have been shown to be related to churn (Bolton 1998; Gustafsson et al. 2005). Since churn events are rare and the size of the dataset is modest, having too many risk categories could also result in overparameterizing the model. As Table 1 shows, controllable churn is the rarest of the three risks we observe, so subdividing that risk even further would not necessarily generate posterior distributions that are any more informative about the distinct components of that risk. In practice, if managers observe many distinct reasons for churn and wish to pool some of them together, the appropriate classification scheme is determined by how similar the reasons are within category (i.e., can reasons in the same category be influenced by the same firm actions?), how distinct they are across categories (i.e., do different firm actions influence different risk categories?), and how prevalent they are in the dataset. Independent of the specific classification scheme employed, the implications of the damper effect in the presence of multiple reasons for churn will hold. As such, for this particular dataset, we employ these three risk categories to demonstrate this effect and the estimation of incremental DERL.

3.1 Estimation and Validation

Following the model specification from Section 2, we define the risk-specific churn processes as two-parameter Weibull distributions with $S_j(t_{ij}|\theta_{ij}) = 2^{-\left(\frac{t_{ij}}{m_{ij}}\right)^{c_{ij}}}$ and $\theta_{ij} = \left[m_{ij}, c_{ij}\right]$, where m_{ij} is customer i's latent median time-to-churn 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 weight. The median parameterization allows for a more intuitive interpretation of coefficients when m_{ij} is regressed on covariates (it is, essentially, a quantile regression), and it is more computationally efficient than using the mean (which depends inconveniently on gamma functions, making it hard to reparameterize the distribution). The Weibull reduces to the exponential distribution when $c_{ij} = 1$. Although the exponential distribution and its discrete-time analog, the geometric distribution, have been employed in previous analyses of customer churn, its "memoryless" property induces a constant hazard rate, and consequently constant relative propensities to churn for each risk throughout the course of a customer relationship. As we will see, the probability of churning for a particular risk changes with a customer's tenure, making the memoryless assumption unreasonable. 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 time-to-churn from a single risk if that risk were the only possible reason for churn. Customers with high m_{ij} are less likely to churn *from risk j* early in the relationship, but some customers might be more likely to churn early when an $m_{ij'}$ is lower for a different risk j'.

The estimated marginal posterior distributions of the model parameters are the empirical distributions of samples from the joint posterior distribution in Equation 7. The data likelihood term $L(\cdot)$ is defined in

Equation 2, using the Weibull definitions of $S_j(\cdot)$ from the last paragraph. The prior on $\log(\theta_i)$ is defined in Equation 6. Therefore, the coefficient matrix Δ is interpreted as the marginal effect of demographic profile x_i on the prior means of $\log(m_{ij})$ and $\log(c_{ij})$ for all individuals and all risks, and Σ is the covariance of the $\log(m_{ij})$ and $\log(c_{ij})$ across customers. The hyperpriors are as described in Section 2. We collect the samples using the MCMC algorithm, whose steps are enumerated in Appendix A.

We ran multiple versions of the model by varying the number of risks (1 or 3), and whether or not we included geodemographic 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 explored the inclusion of time-varying macroeconomic covariates, but found that increasing the dimensionality of the parameter space in this way substantially increased the computational demand of the analysis without improving model fit or predictive performance. Additionally, in our illustrative exercise we will be projecting deep into the future, maintaining parameter estimates that were generated from data collected during a short observation window, for a firm with moderately low monthly churn probabilities. Hence, the specific interpretations or forecasts that depend on model parameters will depend on our distributional assumptions. If a longer observation period were available, allowing us to lengthen our calibration period and have fewer censored observations, we might get different estimates for the posterior distributions of the parameters. We note, though, that such changes would not alter the fundamental concepts of the damper effect in the presence of multiple reasons for which customers churn and the use of incremental DERL as a measure of the return on marketing investment.

3.1.1 Missing Data

One important question is how to handle the 23% of customers who do not report a reason for churn. The answer depends on the assumptions that we make about the randomness of the missing data patterns. Rubin (1976) defines two fundamental properties of missing data: "observed at random" (OAR) and "missing at random" (MAR). Put simply, OAR requires that the values of either the observed data (e.g., measurements or covariates) or missing data do not affect the conditional probability that particular data elements are missing. However, MAR allows for the pattern of missingness to depend on observed data, but not on the values of the missing data themselves. If both MAR and OAR hold, data is considered "missing completely at random" (MCAR), and when MAR does not hold at all, missing data are "not missing at random" (NMAR).

These definitions are useful because if the missing data are either MCAR or MAR, then the missing data process is "ignorable" (Gelman et al. 2003, sec. 21.1), and we do not need to model the selection process explicitly. Ignorability occurs when the parameters that determine the likelihood of the missing data pattern are independent of the parameters on the likelihood of the values of the data, whether missing or not. The independence of parameters means that the likelihood of the missing data pattern can be factored out from the rest of the posterior distribution; the missing data pattern does not influence the posterior distributions of any of the other parameters. But ignorability does not mean that we can simply drop all of the observations for which some data is missing. Dropping incomplete records is permitted only under MCAR, because only under MCAR are all observations equally likely to be dropped. If we were to drop incomplete cases under MAR, the observed dataset might not be representative of the population of customers, because groups with some patterns of observed data could be more likely to be included than others. The typical remedy is through multiple imputation (Little and Rubin 2002, ch. 10). The idea is to "rebalance" the dataset by including all of the records, and treating the missing data as unknown parameters. We then marginalize out the missing data by repeatedly simulating from the conditional posterior distributions of the missing data, given the observed data and the other model parameters. This process is easily incorporated into MCMC algorithms, as we show in Appendix A.

In our dataset, we cannot hold the MCAR assumption to be tenable, because it is possible that customers of certain elapsed tenures or geodemographic profiles could be more likely to have their reasons for churn recorded than others. Since OAR, and thus MCAR, do not apply, simply dropping incomplete records would be inappropriate. Under the MAR assumption, the missing data likelihood can depend on the observed data, but since it is still ignorable, we can adjust for missing data through the multiple imputation step in Appendix A. We readily admit that it is possible that the missing data are not ignorable (which would be the case under NMAR). This would be the case if the reason for churn could influence the likelihood of observing that reason, even after controlling for the observed data. However, we have no prior information to suggest that this is true, and if it were, we lack the additional information that would be necessary to build an explicit model of the missing data process. We therefore retain the MAR assumption as one that is highly probable, and estimate the model parameters accordingly.

3.1.2 Model Performance

Our model validation exercises focus on two managerially relevant quantities: retention rates and risk proportions. The retention rate at time t is $S(t)/S(t-1)$, which is 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 appropriateness of our model, 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 and are equivalent to the stochastic duration-based customer retention models that have appeared in the literature, providing 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 is not surprising that the models without geodemographic 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 presented in the Online Appendix.

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, compared to 2.8% and 2.9%, for the calibration and cross-sectional holdout samples, respectively, for the posterior predictive mean of that distribution from the competing risk model.

Figure 4 shows the actual, and the 95% HPD contours for the posterior predictive distributions, for the proportion of customers who, conditional on churning after a given tenure, churn for each of the three possible risks. Any point within the triangle represents a trivariate vector of probabilities, with the colors of the gridlines matching the color of the axis to which they correspond. The numerals within the triangle indicate the number of months that have elapsed since acquisition and their locations represent the *observed* vector of the proportions of customers of a particular tenure who churn due to a particular risk. The contour lines around each number indicate the 95% HPD region (based on a smoothed kernel density of posterior predictive draws) of the posterior predictive density for the probability vector. In most cases, the locations of the numerals (the observed churn proportions) are reasonably well enclosed by the HPD regions, as can be seen most clearly in the "older" cohorts (11-15), providing some confidence that the model is well-calibrated for predicting how the reason-for-churn proportions evolve over time. We see that customers who churn early in their tenure are most likely to churn from the Nonpay-Abuse risk, but that as time elapses, the churn from Nonpay-Abuse becomes less likely, and churn from uncontrollable reasons becomes more so. Eventually, we see more customers churn from the controllable risk, but this risk is not as prevalent early in the relationship. This pattern mirrors the observed churn of individual cohorts shown in the online appendix. The churn due to nonpay-abuse peaks earlier into a cohort's tenure, and churn from controllable and uncontrollable risks increases later in the relationship. This result is consistent with a "sorting effect" (Fader and Hardie 2010); as the nonpay-abuse-prone customers churn early on, their relative proportion in the population falls, and customers who are likely to churn for other reasons remain. This sorting effect, which had previously not been applied to examine why customers churn, plays a pivotal role in our derivation of incremental DERL, as the firm's efforts to retain customers may deliver a "bigger bang for the buck" when it is more likely that a customer will churn due to a controllable reason compared to other reasons. Note that given 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).

3.2 Interpretation of parameter estimates

Next, we consider posterior distributions of the baseline median lifetimes for both the single-risk and multiple-risk cases, and how these median lifetimes vary across geodemographic groups.[6] Table 2 summarizes the posterior means and 95% HPD intervals of the percentage differences in median lifetimes that are attributable to demographic factors. The first two lines are the median times to churn from each of the three risks (the Controllable, Uncontrollable and Nonpay-Abuse columns) from the competing risk model, and the aggregate median lifetimes (the All column). The reported median lifetimes under the model with

[6] Summary tables of the estimated posterior distributions of Δ and Σ are available in the Online Appendix.

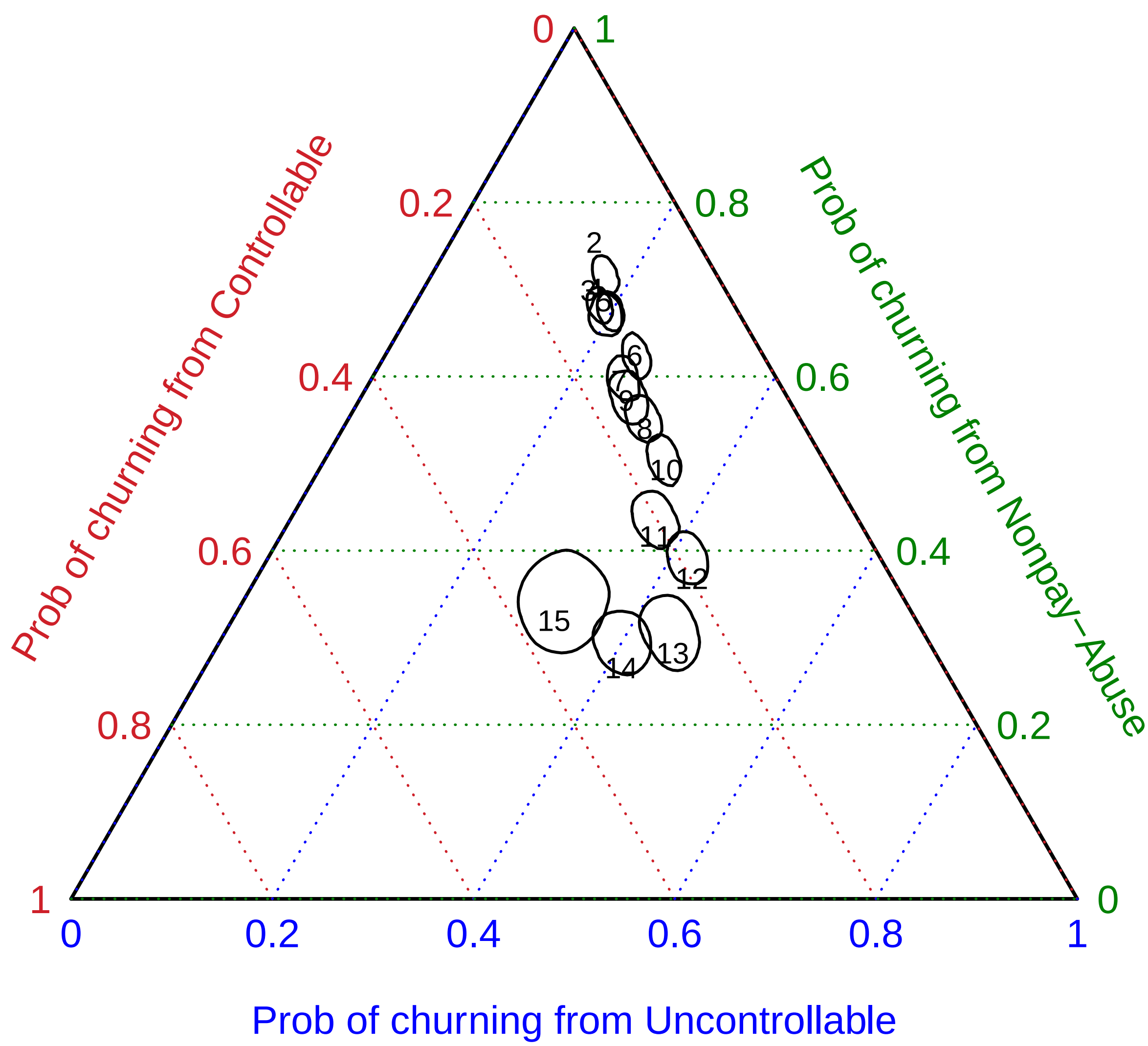

Figure 4: Observed and posterior HPD contours for the relative proportions of reasons for churn for customers who have survived a certain number of months. Any point within the triangle represents a trivariate vector of probabilities. The arrows along each axis indicate which set of gridlines correspond 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.

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 from the controllable 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 geodemographic 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 nonpay-abuse, which also happens to the most prevalent cause of churn for all groups.

We can see these variations more clearly in Figure 5. Each panel represents the posterior means of the hazard rate functions for 24 months, broken down by geodemographic 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 geodemographic 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 nonpay-abuse is relatively high at the start of a customer's relationship, but declines sharply as time passes. These differences, both across geodemographic 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 geodemographic clusters, and at different points in the relationship can have varying effects in the long-term value of the customer base.

3.3 Posterior Incremental DERL

We now revisit the concept of the damper effect and its implications for incremental DERL, initially introduced in Section 2.2, based on the model results from our empirical analysis. Figure 6 plots how DERL, discounted back to today, for a hypothetical customer in each of five geodemographic groups,

	Controllable			Uncontrollable			Nonpay-Abuse			All		
	HPD Lower	**Mean**	HPD Upper	HPD Lower	**Mean**	HPD Upper	HPD Lower	**Mean**	HPD Upper	HPD Lower	**Mean**	HPD Upper
Models with no demographics												
Median lifetime	103.9	**119.0**	140.0	57.3	**61.1**	65.2	45.1	**47.4**	49.7	35.1	**36.6**	38.0
Models with demographics												
Baseline median lifetime (months)	92.3	**106.1**	121.4	52.7	**56.5**	60.3	40.1	**42.0**	43.9	31.5	**32.8**	34.0
Percentage differences between groups												
Loc - Town/Rural : Urban	0.6	**0.8**	0.9	0.8	**0.9**	1.0	1.1	**1.1**	1.2	0.9	**1.0**	1.1
Loc - Suburban : Urban	-6.5	**-3.7**	-1.1	-0.8	**1.0**	3.5	-5.1	**-3.2**	-1.4	-3.6	**-2.1**	-0.4
Income - Middle : Low	-12.6	**-10.7**	-9.1	-14.6	**-13.4**	-12.3	-16.8	**-15.8**	-14.9	-14.8	**-14.1**	-13.1
Income - High : Low	13.6	**16.7**	19.2	18.2	**20.1**	22.0	23.8	**25.5**	27.5	20.4	**22.1**	23.6
Age - Middle : Low	9.7	**12.2**	15.3	8.2	**10.1**	12.1	14.9	**16.5**	18.6	12.0	**13.7**	15.3
Age - High : Low	-4.6	**-3.2**	-2.0	-2.2	**-1.3**	-0.2	-4.5	**-3.7**	-2.8	-3.7	**-2.9**	-2.1
Kids - Yes : No	-0.2	**2.7**	5.6	-4.6	**-2.4**	-0.5	-0.4	**1.6**	3.5	-1.1	**0.6**	2.2
Home - Rent : Own	-12.2	**-9.3**	-6.6	-16.5	**-14.5**	-12.9	-16.4	**-14.9**	-13.2	-15.0	**-13.7**	-12.1
Job - White Collar : Blue Collar	0.5	**0.6**	0.7	0.4	**0.5**	0.6	0.7	**0.8**	0.9	0.6	**0.7**	0.7
Job - Prof/Mgmt : Blue Collar	8.7	**11.3**	13.9	14.2	**16.0**	17.8	16.6	**18.2**	19.8	14.9	**16.2**	17.6
Job - Retired : Blue Collar	-3.4	**-2.5**	-1.5	-1.7	**-1.0**	-0.3	-3.5	**-2.9**	-2.2	-2.9	**-2.2**	-1.7
Edu - HS Grad : Some HS	-6.2	**-5.2**	-4.3	-7.5	**-6.9**	-6.2	-8.5	**-7.9**	-7.4	-7.5	**-7.1**	-6.5
Edu - Some Coll : Some HS	-9.4	**-8.0**	-6.7	-8.5	**-7.5**	-6.5	-11.6	**-10.8**	-10.0	-10.0	**-9.3**	-8.5
Edu - Coll Grad : Some HS	7.5	**9.4**	11.7	6.4	**7.8**	9.4	11.5	**12.7**	14.3	9.3	**10.6**	11.7
Edu - Grad Plus : Some HS	5.0	**6.6**	8.4	8.7	**9.8**	11.0	9.7	**10.7**	11.8	8.7	**9.6**	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.

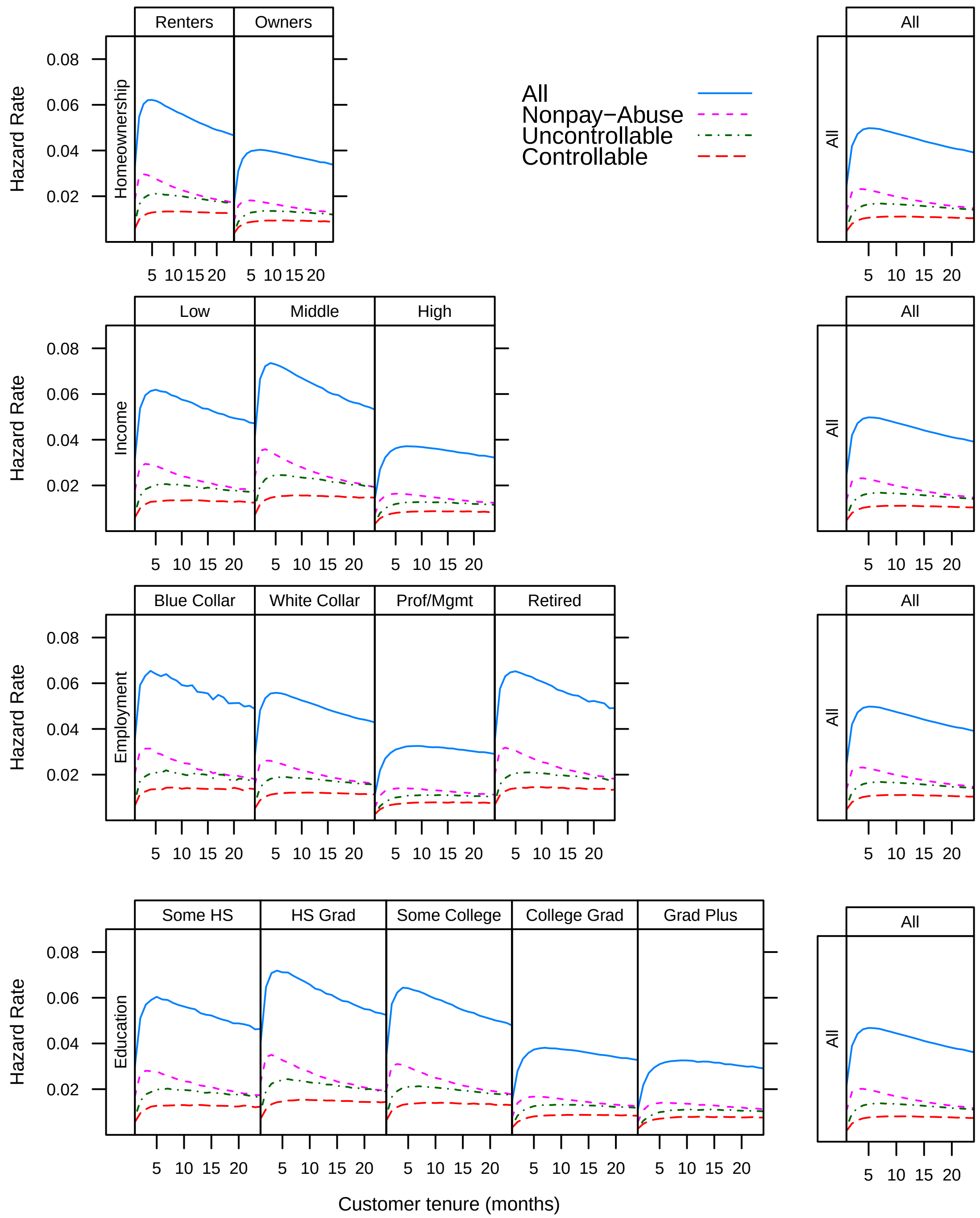

Figure 5: Posterior mean hazard rate functions, by demographic group, for each of the competing risks, and all risks combined. Panels in each row correspond to a demographic factor (Homeownership, Income Level, Employment Type and Education Level). The "All" panels on the right correspond to the entire population, and are replicated across rows to facilitate comparisons.

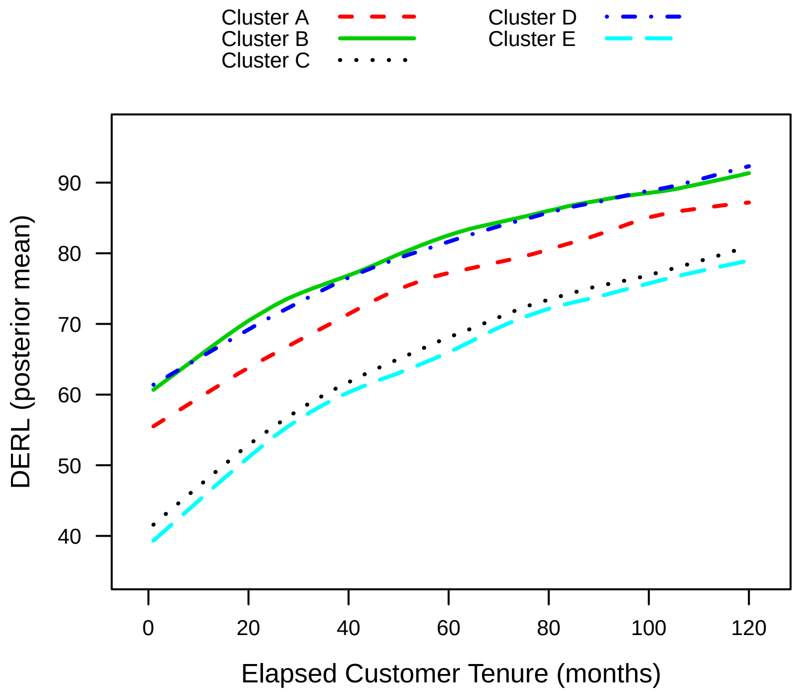

Figure 6: Posterior expected values for discounted expected residual lifetime (DERL) for five geodemographic clusters. Churn propensities for customers with long tenure with the company are more likely to be lower than more recent customers, and so we expect their lifetimes to be longer. The descriptions of the clusters are in Table 3

would change for different durations of service T. Descriptions of these clusters are in Table 3. Customers who have a longer tenure with the firm are less likely to have high churn rates, as reflected in the posterior estimates of θ_i, leading to an increasing pattern in DERL as T increases.

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 9 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 propensity 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 geodemographic 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., service improvements or bill collection). Depending on how far the manager pulls lever J, he delays the median time-to-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 10 to something higher. This difference in DERLs depends on θ_i, which is unobserved. Integrating the difference between the baseline DERL and the "μ_j-enhanced" DERL over the posterior distribution of θ_i given T yields the expected increase a customer's discounted remaining lifetime and can be used by a manager as an upper bound on expenditures to pull lever j. With θ_i integrated out, DERL depends on T, which is observable to a manager, and μ_j, which is related to his "pulling the levers". This expected difference between the baseline DERL and the μ_j-enhanced DERL is the expected incremental DERL.

Figure 7 illustrates the relationships among T, μ_j and expected incremental DERL for customers in the same five geodemographic clusters that were described in Table 3 and used for Figure 6. Each row of panels corresponds to geo-demographic cluster, and each column of panels represents one of the three risks in our empirical analysis. 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 controllable reasons has relatively little affect on DERL. Keeping Equation 9 and Figure 5 in mind, this should not come as a surprise because the controllable risk is the least prevalent risk and so the damper ratio from Equation 9 is closer to zero. This is a critical realization for a firm for which the controllable risk is the only one on which many marketing efforts may have any effect. Depending on the how much it costs to the extend the latent lifetime from the controllable risk, 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.

Looking at the controllable and nonpay-abuse risks, we observe some "backward bending" patterns in the surfaces in Figure 7. 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 6. These findings are

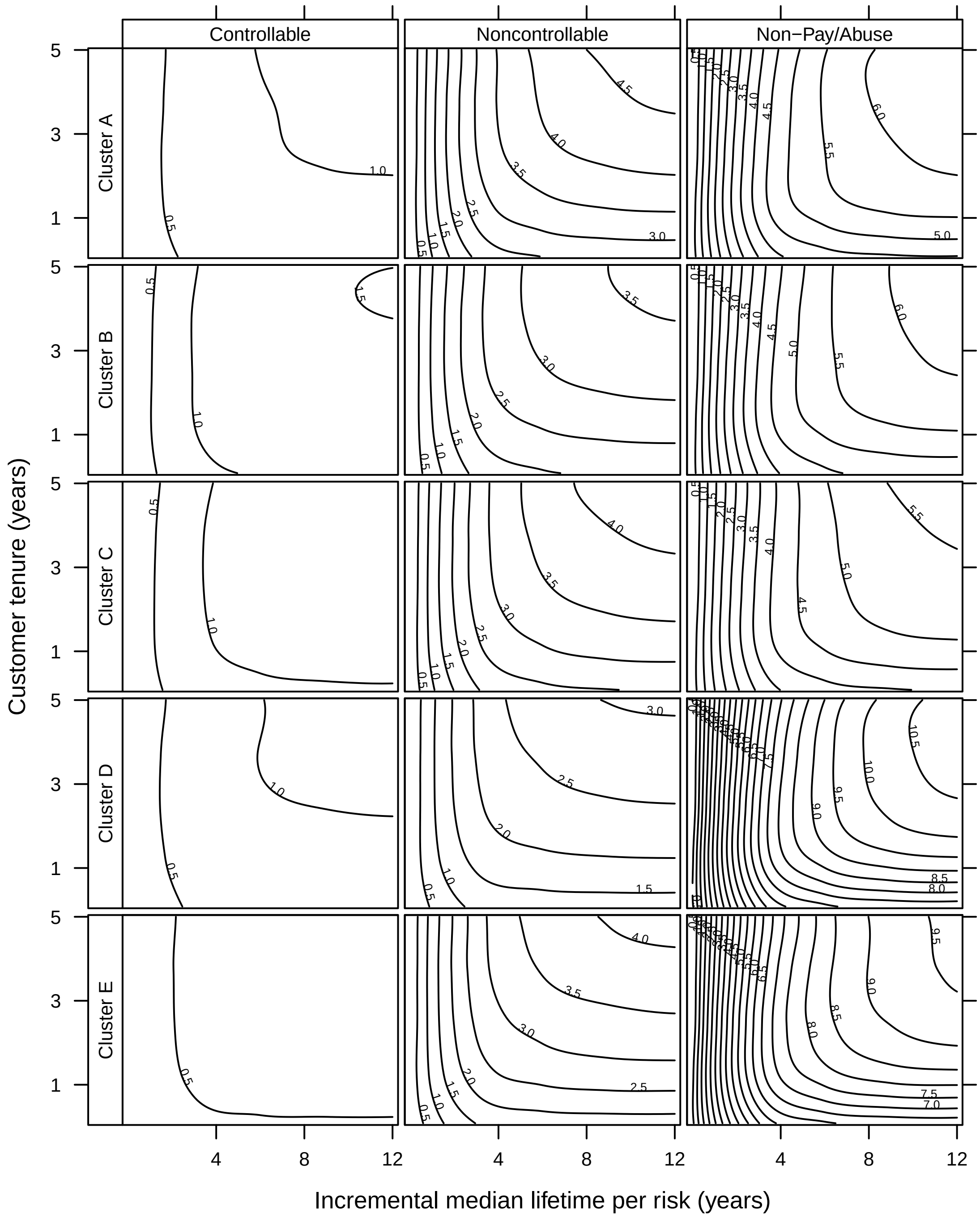

Figure 7: Contour lines representing the incremental discounted expected residual lifetime (DERL) that a firm can expect by delaying churn that is due to a specific risk. The x-axis represents the additional number of years added to the remaining median risk-specific lifetime for each risk. The y-axis represents the number of years of the tenure of the existing customer relationship. DERL was computed using an annual discount rate of 15 percent.

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. This leads to an initial decline in customer retention, as can be seen in the posterior retention curves in the Online Appendix, 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 whom 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 geodemographic clusters and across the risks. First, consider the impact of delaying controllable churn. We see that the incremental DERL "surface" is low and flat across the population. As such, no matter how much extra μ_j the firm adds to a customer, it recovers only fractions of additional DERL units. If the firm were to decide to invest in delaying controllable churn, it may be best to do so for customers in clusters B and C. We see the closest incremental DERL contours in these clusters, suggesting that they are the most responsive. Within these clusters, it may be best to invest in "older" customers.

In our empirical application, the controllable risk is the least prevalent risk and hence the one for which changes to the median remaining lifetime will have little effect on DERL. However, 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 nonpay-abuse, such as by increasing its bill collection efforts. As we see from Figure 7, doing so could be quite fruitful. In comparison to the controllable risk, the impact of increasing μ_j climbs more steeply, particularly for customers with a longer elapsed tenure. As seen in Figure 5, the tendency to churn due to nonpay-abuse is high early in customers' relationships. When a customer 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, leading the contour lines to start turning 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. As was the case with the controllable risk, we again see variation in the incremental DERL across clusters for nonpay-abuse, with customers in clusters D and E potentially being affected the most by an increase in μ_j.

Hypothetically, suppose that the firm could in some way influence uncontrollable churn. 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 clusters B and D, which are comprised of homeowners in less urban areas. Again we see how delaying churn might have different effects on the incremental 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." Reducing churn from the most prevalent risk, as shown in Equation 9, has the largest potential effect on increasing customers' lifetimes. Figure 7 provides managers with a sense for the potential impact that their actions may have on extending customers' remaining lifetimes, based on the firm's historical data.

There is, of course, some uncertainty in the estimates of incremental DERL. In Figure 8, we give a sense of the posterior variation of these estimates through the case of a customer from Cluster D, with an elapsed tenure of 45 months. As in Figure 7, the x-axis is μ_j, the incremental lifetime added to the median time-to-churn for each of the three risks. The y-axis is the incremental DERL that the firm can expect to accrue from adding that additional risk-specific lifetime. The solid line shows the posterior mean of incremental DERL and corresponds with the "height" of the surface in Figure 7. The dashed line represents the 95% quantile. The intensity of the gray scale represents the amount of posterior mass, so darker areas are more likely to occur than lighter areas. For this customer, we see a high probability of there being very little incremental DERL, but the tail of this posterior distribution is long. Though there is a high potential upside for a customer's incremental DERL, managers should be cognizant of the uncertainty in how much return they can expect from extending the risk-specific remaining lifetimes.

4 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 research describes a novel approach to applying competing risk models to an important managerial setting and offers insights for managers that extant retention models do not.

As we illustrate, disentangling the likelihood of churn due to different reasons provides managers

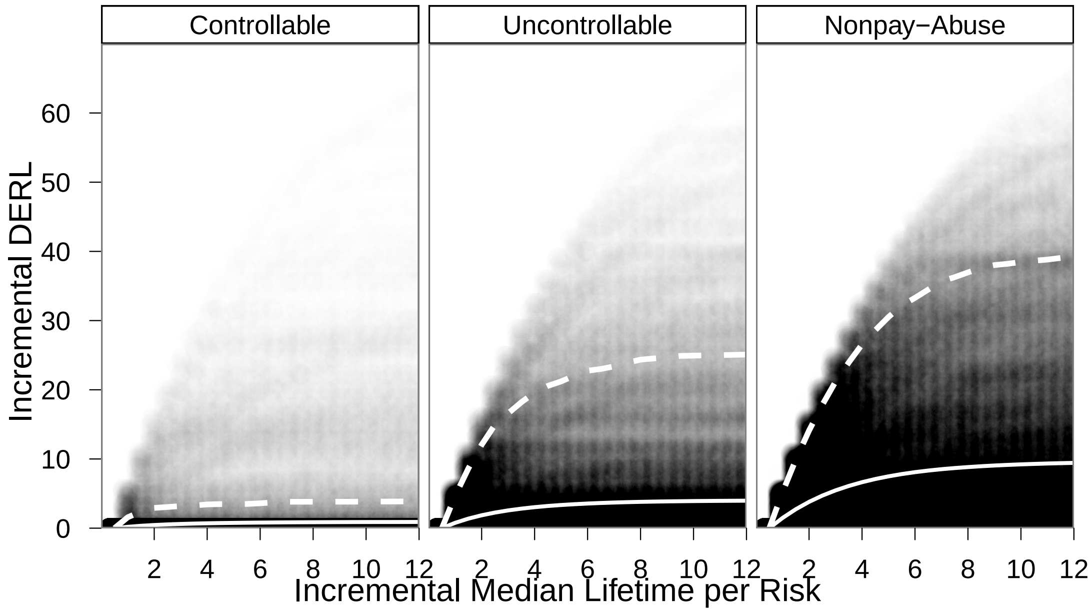

Figure 8: The posterior distribution of incremental DERL for a customer in Cluster A, whose duration of relationship with the firm is eight years.

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 damper effect (Equation 9) that arises due to the presence of alternative reasons for churn, including uncontrollable reasons. As the 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 used to prioritize spending on customers or establish upper bounds on such investments, based on the expected returns (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. Firms relying on raw retention metrics may erroneously conclude that their retention marketing activities have little effect when their actions are actually quite effective, but only for a subset of customers who are likely to discard service for a reason that the firm can influence. Moreover, raw retention metrics may mask shifts in the balance between controllable and uncontrollable churn, which may be informative from a managerial perspective. Much like the claim that 100% loyalty is an important managerial objective (Reichheld and Sasser 1990; Jones and Sasser 1995) has been challenged because it is 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 (Seetharaman and Chintagunta 2003), though it increases the computational burden of the analysis. This would permit the inclusion of information on current economic conditions, marketing activities, service quality or customer touchpoints with the service provider when available. Such data would allow for an investigation of how multiple potential triggers, such as price hikes *and* service problems, may accumulate and jointly contribute to a customer's disutility with a service provider, ultimately leading to churn when customers exceed a level of tolerance. 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. Similarly, detailed data on the ownership and usage of individuals services may provide additional insight into the ways in which to manage customer retention, as well as cross-selling and up-selling (Li et al. 2005; Schweidel et al. 2011). We believe that incorporating such information (when available) 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 damper effect that we have discussed, allowing stakeholders to more clearly evaluate the impact of the service provider's investments in marketing and service improvements. The same holds true within an organization, as marketers need to demonstrate the efficacy of their actions. 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).

Although 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. For the purposes of our illustration and to keep the analysis tractable, we assumed that the marketing "levers" affect only a single cause of churn in our illustrations. But if a particular marketing activity could delay more than one cause of churn, the firm can benefit from multiple amounts of incremental DERL. Firms should take this into account when comparing the expected benefit of retention marketing activities to the cost.

For a firm to maintain its market share, it will have to replace the customers it loses (regardless of the reason for which they 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 same activities that contribute to increased retention may also influence the acquisition process. Since marketing actions and service improvements can influence both the acquisition and retention processes simultaneously, firms should evaluate these processes jointly (Schweidel et al. 2008b). When firms have a fixed marketing budget, the competing risk framework offers firms some guidance as to how to allocate and deploy marketing resources. Because the reasons for which a customer is likely to churn may evolve as time elapsed since acquisition, an "optimal" balance between acquisition and retention activities may depend on current customers' tenures to date (Reinartz et al. 2005). Therefore, marketing efforts and improvements to service quality, for example, may play a larger role in the acquisition process than in the retention process. Of course, these tradeoffs depend on the firm and the context.

Appendices

A MCMC Algorithm for Model Estimation

We use Gibbs sampling to estimate marginal posterior distributions of θ_i (for $i = 1 \dots N$), 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 $\theta_i|\cdot$

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

$$\pi(\theta_i|\Delta, x_i, \Sigma) \propto |\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\theta_i - \Delta x_i)'\Sigma^{-1}(\theta_i - \Delta x_i)\right] \tag{12}$$

The general form of the log conditional posterior (ignoring normalizing constants) is the log of data likelihood in Equation 2, plus the log of the prior distribution in Equation 12.

$$\log \pi(\theta_i|t_i, j_i, \Delta, x_i, \Sigma) = \log \pi(\theta_i|\Delta, x_i, \Sigma) + \begin{cases} \log\left[S_j(t_i - 1|\theta_{ij}) - S_j(t_i|\theta_{ij})\right] + \log S(t_i - 1|\theta_i) & \text{if } d_i = 1 \\ \log S(T_i|\theta_i) & \text{if } d_i = 0 \end{cases} \tag{13}$$

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

$$S_j(t_i|\theta_{ij}) = 2^{-\left(\frac{t_i}{m_{ij}}\right)^{c_{ij}}}, \tag{14}$$

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 13 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 $\Delta|\cdot$

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

$$\pi\left(\operatorname{vec}\Delta|\Delta_0, \Omega, \Sigma, x\right) \propto |\Omega \otimes \Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\operatorname{vec}\left(\Delta - \Delta_0\right)'\left(\Omega^{-1} \otimes \Sigma^{-1}\right)\operatorname{vec}\left(\Delta - \Delta_0\right)\right] \tag{15}$$

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

$$\pi(\Theta|X, \Delta, \Sigma) \propto |\Sigma|^{-\frac{N}{2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{N}(\theta_i - \Delta x_i)'\Sigma^{-1}(\theta_i - \Delta x_i)\right] \tag{16}$$

Equation 16 depends on Δ, not $\operatorname{vec}\Delta$. So, expressing Δx_i as $I_{rJ}\Delta x_i$ (I_{rJ} is the $rJ \times rJ$ identity matrix), and applying the identity $\operatorname{vec}(ABC) = (C' \otimes A) \operatorname{vec} B$, we write the joint prior distribution for Θ as

$$\pi(\Theta|X, \Delta, \Sigma) \propto |\Sigma|^{-\frac{N}{2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{N}\left(\theta_i - \left(x_i' \otimes I_{m(k+g)}\right)\operatorname{vec}\Delta\right)' \Sigma^{-1}\left(\theta_i - \left(x_i' \otimes I_{m(k+g)}\right)\operatorname{vec}\Delta\right)\right] \tag{17}$$

No other terms in the joint posterior distribution involve Δ, so we can get the conditional posterior distribution for $\operatorname{vec}\Delta$ by multiplying Equations (15) and (17) together, and simplifying the result by "completing the square."

$$\begin{aligned}
\pi\left(\operatorname{vec}\Delta|\theta, \Omega, \Sigma, \Delta_0\right) \propto \exp\Bigg[&-\frac{1}{2}\operatorname{vec}\left(\Delta - \Delta_0\right)'\left(\Omega^{-1} \otimes \Sigma^{-1}\right)\operatorname{vec}\left(\Delta - \Delta_0\right) \\
&+\sum_{i=1}^{N}\left(\theta_i - \left(x_i' \otimes I_{m(k+g)}\right)\operatorname{vec}\Delta\right)' \Sigma^{-1}\left(\theta_i - \left(x_i' \otimes I_{m(k+g)}\right)\operatorname{vec}\Delta\right)\Bigg] \\
= \exp\Bigg[-\frac{1}{2}\Bigg(&\operatorname{vec}\Delta'\left(\left(\Omega^{-1} \otimes \Sigma^{-1}\right) + \sum_{i=1}^{N}\left(x_i \otimes I_{m(k+g)}\right)\Sigma^{-1}\left(x_i' \otimes I_{m(k+g)}\right)\right)\operatorname{vec}\Delta \\
&-2\operatorname{vec}\Delta'\left(\left(\Omega^{-1} \otimes \Sigma^{-1}\right)\operatorname{vec}\Delta_0 + \sum_{i=1}^{N}\left(x_i \otimes I_{m(k+g)}\right)\Sigma^{-1}\theta_i\right) + C\Bigg)\Bigg] \\
= \exp\Bigg[-\frac{1}{2}\Bigg(&\operatorname{vec}\Delta'\left(\left(\Omega^{-1} \otimes \Sigma^{-1}\right) + \left(\sum_{i=1}^{N} x_i x_i' \otimes \Sigma^{-1}\right)\right)\operatorname{vec}\Delta \\
&-2\operatorname{vec}\Delta'\left(\left(\Omega^{-1} \otimes \Sigma^{-1}\right)\operatorname{vec}\Delta_0 + \sum_{i=1}^{N}\left(x_i \otimes \Sigma^{-1}\theta_i\right)\right) + C\Bigg)\Bigg]
\end{aligned} \tag{18}$$

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

Equation 18 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

$$\begin{aligned}
\operatorname{cov}(\operatorname{vec}\Delta|\cdot) &= \left(\left(\Omega^{-1} + X'X\right) \otimes \Sigma^{-1}\right)^{-1} \\
&= \left(\Omega^{-1} + X'X\right)^{-1} \otimes \Sigma
\end{aligned} \tag{19}$$

The conditional posterior mean is also readily extracted from Equation 18, and after some tedious and mechanical manipulations can be simplified as

$$\begin{aligned}
E(\operatorname{vec}\Delta|\cdot) &= \left(\left(\Omega^{-1} + X'X\right)^{-1} \otimes \Sigma\right)\left(\left(\Omega^{-1} \otimes \Sigma^{-1}\right)\operatorname{vec}\Delta_0 + \sum_{i=1}^{N}\left(x_i \otimes \Sigma^{-1}\theta_i\right)\right) \\
&= \operatorname{vec}\left(\left(\Delta_0\Omega^{-1} + \Theta'X\right)\left(\Omega^{-1} + X'X\right)^{-1}\right)
\end{aligned} \tag{20}$$

As Rossi et al. (2005) discuss, one might be tempted to simulate $\mathrm{vec}\,\Delta$ 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 $\mathrm{vec}\,\Delta$ (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. We replicate those steps here.

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 $W = \binom{X}{\Lambda'}$ 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 $\mathrm{vec}\,\Delta$ as

$$\begin{aligned}
\mathrm{cov}\left(\mathrm{vec}\,\Delta|\cdot\right) &= \left(X'X + \Omega^{-1}\right)^{-1} \otimes \Sigma \\
&= \left(W'W\right)^{-1} \otimes \left(LL'\right)^{-1} \\
&= \left(R_{W'W}R'_{W'W}\right)^{-1} \otimes \left(LL'\right)^{-1} \\
&= \left(R^{-1}_{W'W} \otimes L^{-1}\right)' \left(R^{-1}_{W'W} \otimes L^{-1}\right)
\end{aligned} \tag{21}$$

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 $\mathrm{vec}\,\Delta$ by multiplying Ψ by the Cholesky root of $\mathrm{vec}\,\Delta$'s posterior covariance, and then adding its posterior mean. Some more mechanical manipulations get us to a posterior draw of Δ.

$$\begin{aligned}
\mathrm{vec}\,\Delta &= \mathrm{vec}\left(\left(\Delta_0\Omega^{-1} + \Theta'X\right)\left(\Omega^{-1} + X'X\right)^{-1}\right) + \left(R^{-1}_{W'W} \otimes L^{-1}\right)' \mathrm{vec}\,\Psi \\
&= \mathrm{vec}\left(\left(\Delta_0\Omega^{-1} + \Theta'X\right)\left(\Omega^{-1} + X'X\right)^{-1}\right) + \mathrm{vec}\left(L'^{-1}\Psi R^{-1}_{W'W}\right) \\
\Delta &= \left(\Delta_0\Omega^{-1} + \Theta'X\right)\left(\Omega^{-1} + X'X\right)^{-1} + L'^{-1}\Psi R^{-1}_{W'W}
\end{aligned} \tag{22}$$

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 $\Sigma|\cdot$

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

$$\pi\left(\Sigma|\nu, A\right) \propto |\Sigma|^{-\frac{\nu+m(k+g)+1}{2}} \exp\left[-\frac{1}{2}\operatorname{tr}\left(A\Sigma^{-1}\right)\right] \tag{23}$$

The conditional posterior distribution for Σ is found multiplying Equations (15), (16), and (23). The result is an inverse Wishart distribution with $\nu + p + N$ degrees of freedom and a location parameter $A + \left[\left(\Delta - \Delta_0\right)\Omega^{-1}\left(\Delta - \Delta_0\right)'\right] + \left(\Theta - x\Delta'\right)'\left(\Theta - x\Delta'\right)$. 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 2, 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 9

Starting from Equation 8, the marginal effect of a single risk-specific parameter θ_j on ECLV is

$$\frac{\partial ECLV}{\partial \theta_j} = \sum_{t=1}^{\infty} \delta^t \frac{\partial S(t|\theta)}{\partial \theta_i} \tag{24}$$

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

$$S(t|\theta) = \exp\left[\log S_j(t|\theta_j) + \sum_{k \neq j} \log S_k(t|\theta_k)\right]$$
$$\frac{\partial S(t|\theta)}{\partial \theta_j} = S(t|\theta) \frac{1}{S_j(t|\theta_j)} \frac{\partial S_j(t|\theta_j)}{\partial \theta_j} \tag{25}$$

Then substitute Equation 25 into Equation 24.

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,

$$S(t|m, c) = 2^{-\left(\frac{t}{m}\right)^c} \tag{26}$$

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

$$S(t|m, c, t > T) = 2^{-\left(\frac{t^c - T^c}{m^c}\right)} \tag{27}$$

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

$$m^* = (m^c + T^c)^{\frac{1}{c}} \tag{28}$$

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^*,

$$m^* + \mu = \left(m_\mu^c + T^c\right)^{\frac{1}{c}}$$
$$m_\mu = \left[(m^* + \mu)^c - T^c\right)^{\frac{1}{c}} \tag{29}$$

Substituting 28 into 29 yields

$$m_\mu = \left(\left[(m^c + T^c)^{\frac{1}{c}} + \mu \right]^c - T^c \right)^{\frac{1}{c}} \tag{30}$$

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

References

Jaap H Abbring and Gerard J van den Berg. The Identifiability of the Mixed Proportional Hazards Competing Risks Model. *Journal of the Royal Statistical Society, Series B*, 65(3):701–710, 2003.

Jan Beyersmann, Aurelien Latouche, Anika Buchholz, and Martin Schumacher. Simulating Competing Risks Data in Survival Analysis. *Statistics in Medicine*, 28:956–971, 2009.

Svetlana Bogomolova and Jenni Romaniuk. Brand Defection in a Business-to-Business Financial Service. *Journal of Business Research*, 62:291–296, 2009.

Ruth N Bolton. 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, 1998.

Sharad Borle, Siddharth S Singh, and Dipak C Jain. Customer Lifetime Value Measurement. *Management Science*, 54(1):100–112, jan 2008.

Pradeep K Chintagunta. Inertia and Variety Seeking in a Model of Brand-Purchase Timing. *Marketing Science*, 17(3):253–270, jan 1998.

David Clayton and Jack Cuzick. Multivariate Generalizations of the Proportional Hazards Model. *Journal of the Royal Statistical Society. Series A (General)*, 148(2):82–117, 1985.

Andrew S C Ehrenberg, Gerald J Goodhardt, and T Patrick Barwise. Double Jeopardy Revisited. *Journal of Marketing*, 52(2):82–91, 1990.

Peter S Fader and Bruce G S Hardie. Probability Models for Customer Base Analysis. *Journal of Interactive Marketing*, 23:61–69, 2009.

Peter S Fader and Bruce G S Hardie. Customer-Base Valuation in a Contractual Setting: The Perils of Ignoring Heterogeneity. *Marketing Science*, 29(1):85–93, jan 2010.

Peter S Fader, Bruce G S Hardie, and Yuzhou Liu. Customer Retention Dynamics in a Contractual Setting: The Paradox of Increasing Loyalty. Working paper, Wharton School, University of Pennsylvania, sep 2009.

Drew Fudenberg and J Miguel Villas-Boas. Behavior-Based Price Discrimination and Customer Recognition. *Economics and Information Systems*, 1, 2007.

Andrew Gelman, Xiao-Li Meng, and Hal Stern. Posterior Predictive Assessment of Model Fitness via Realized Discrepancies. *Statistica Sinica*, 6(4):733–807, 1996.

Andrew Gelman, John B Carlin, Hal S Stern, and Donald B Rubin. *Bayesian Data Analysis*. Chapman and Hall, Boca Raton, Fla., 2003.

Sunil Gupta, Donald R Lehmann, and Jennifer Ames Stuart. Valuing Customers. *Journal of Marketing Research*, 41(1):7–18, 2004.

Anders Gustafsson, Michael D Johnson, and Inger Roos. The Effects of Customer Satisfaction, Relationship Commitment Dimensions, and Triggers on Customer Retention. *Journal of Marketing*, 69:210–218, oct 2005.

James J Heckman and Bo E Honoré. The Identifiability of the Competing Risks Model. *Biometrika*, 76(2): 325–330, jun 1989.

David G Hoel. A Representation of Mortality Data by Competing Risks. *Biometrics*, 28(2):475–488, jun 1972.

John E Hogan, Katherine N Lemon, and Roland T Rust. Customer Equity Management: Charting New Directions for the Future of Marketing. *Journal of Service Research*, 5(1):4–12, aug 2002.

Thomas O Jones and W Earl Sasser. Why Satisfied Customers Defect. *Harvard Business Review*, 73:88–99, nov 1995.

Susan M Keaveney. Customer Switching Behavior in Service Industries: An Exploratory Study. *Journal of Marketing*, 59(2):71–82, apr 1995.

Timothy Keiningham, Terry G Vavra, Lerzan Aksoy, and Henri Wallard. *Loyalty Myths*. John Wiley and Sons, Hoboken, N.J., 2005.

Donald R Lehmann. Metrics for Making Marketing Matter. *Journal of Marketing*, 68:73–75, oct 2004.

Shibo Li, Baohung Sun, and Ronald T Wilcox. Cross-Selling Sequentially Ordered Products: An Application to Consumer Banking Services. *Journal of Marketing Research*, 42(2):233–239, 2005.

Roderick J A Little and Donald B Rubin. *Statistical Analysis with Missing Data*. Wiley-Interscience, Hoboken, N.J., 2002.

Wendy W Moe and Michael Trusov. Measuring the Value of Social Dynamics in Online Product Ratings Forums. *Journal of Marketing Research*, 2010.

Richard L Oliver. Whence Customer Loyalty? *Journal of Marketing*, 63:33–44, 1999.

Don O'Sullivan and Andrew V Abela. Marketing Performance Measurement Ability and Firm Performance. *Journal of Marketing*, 71:79–93, apr 2007.

R L Prentice, J D Kalbfleisch, A V Peterson, N Flournoy, and V T Farewell. The Analysis of Failure Times in the Presence of Competing Risks. *Biometrics*, 34(4):541–554, dec 1978.

H Putter, M Fiocco, and R B Geskus. Tutorial in Biostatistics: Competing Risks and Multi-State Models. *Statistics in Medicine*, 26:2389–2430, 2007.

Priya Raghubir and Sanjiv R Das. The Long and Short of It: Why are Stocks with Shorter Runs Preferred? *Journal of Consumer Research*, 36:964–982, apr 2010.

Frederick F Reichheld and W Earl Sasser. Zero Defections: Quality Comes to Services. *Harvard Business Review*, 68:105–111, sep 1990.

Werner Reinartz, Jacquelyn S Thomas, and V Kumar. Balancing Acquisition and Retention Resources to Maximize Customer Profitability. *Journal of Marketing*, 69:63–79, jan 2005.

Werner J Reinartz and V Kumar. The Impact of Customer Relationship Charactersitics on Profitable Lifetime Duration. *Journal of Marketing*, 67(1):77–99, 2003.

Peter E Rossi, Greg M Allenby, and Robert McCulloch. *Bayesian Statistics and Marketing*. John Wiley and Sons, Chichester, UK, 2005.

Donald B Rubin. Inference and Missing Data. *Biometrika*, 63(3):581–592, dec 1976.

Donald B Rubin. Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician. *Annals of Statistics*, 12(4):1151–1172, 1984.

Roland T Rust, Anthony J Zahorik, and Timothy Keiningham. Return on Quality (ROQ): Making Service Quality Financially Accountable. *Journal of Marketing*, 59(2):58–70, apr 1995.

Roland T Rust, Katherine N Lemon, and Valarie A Zeithaml. Return on Marketing: Using Customer Equity to Focus Marketing Strategy. *Journal of Marketing*, 68(1):109–127, jan 2004.

David A Schweidel, Peter S Fader, and Eric T Bradlow. Understanding Service Retention Within and Across Cohorts Using Limited Information. *Journal of Marketing*, 72:82–94, jan 2008a.

David A Schweidel, Peter S Fader, and Eric T Bradlow. A Bivariate Timing Model of Customer Acquisition and Retention. *Marketing Science*, 27(5):829–843, sep 2008b.

David A Schweidel, Eric T Bradlow, and Peter S Fader. Portfolio Dynamics for Customers of a Multi-Service Provider. *Management Science*, 57(3):829–843, 2011.

P B Seetharaman and Pradeep K Chintagunta. The Proportional Hazard Model for Purchase Timing. *Journal of Business and Economic Statistics*, 21(3):368–382, jul 2003.

Greg Shaffer and Z John Zhang. Pay to Switch or Pay to Stay: Preference-Based Price Discrimination in Markets with Switching Costs. *Journal of Economics and Management Strategy*, 9(3):397–424, 2000.

Greg Shaffer and Z John Zhang. Competitive One-to-One Promotions. *Marketing Science*, 48(9):1143–1160, sep 2002.

Byron Sharp. *How Brands Grow*. Oxford University Press, Melbourne, Australia, 2010.

Raji Srinivasan, Gary L Lilien, and Arvind Rangaswamy. 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, 2008.

Curtis R Taylor. Supplier Surfing: Competition and Consumer Behavior in Subscription Markets. *The RAND Journal of Economics*, 34(2):223–246, 2003.

Rajkumar Venkatesan and V Kumar. A Customer Lifetime Value Framework for Customer Selection and Resource Allocation Strategy. *Journal of Marketing*, 68(4):106–125, oct 2004.

Peter C Verhoef. Understanding the Effect of Customer Relationship Management Efforts on Customer Retention and Customer Share Development. *Journal of Marketing*, 67(4):30–45, oct 2003.

Naufel J Vilcassim and Dipak C Jain. Modeling Purchase-Timing and Brand-Switching Behavior Incorporating Explanatory Variables and Unobserved Heterogeneity. *Journal of Marketing Research*, 28(1):29–41, feb 1991.