21

HAZARD/SURVIVAL MODELS IN MARKETING

PRADEEP K. CHINTAGUNTA
University of Chicago

XIAOJING DONG
Northwestern University

A variable of considerable interest to marketing researchers is time. Time can take on the role of an explanatory variable in many contexts, for example, when one is dealing with data that reflect a temporal trend. At the same time, there are several situations in which time is the dependent variable that the researcher would like to focus on. For example, one might be interested in the duration of a relationship between a service provider (such as an advertising agency) and its client (the firm employing the ad agency) and in studying the effects of various factors that might influence the duration of the relationship. Survival models—also referred to as duration models, hazard models, or failure time models—are uniquely suited to addressing such issues.

Before proceeding with a description of the model itself, it may be worthwhile asking the following: Since time is a continuous variable and since regression methods are very well developed, why not simply use a regression model to study the influence of explanatory variables on the durations of interest? There are two main reasons. The first is referred to as "*censoring*" or, more precisely, for a vast majority of topics of interest, "right censoring" The second is the issue of handling explanatory variables or factors that vary over time, also known as "*time-varying covariates.*"

The censoring issue can be explained as follows. Consider, again, the example of advertising agency relationship durations. Suppose we have a sample of such relationships that were established in the year 1990 and we observed these relationships until the year 2000. For each such relationship, the data would indicate two possibilities: (a) where the relationship terminates between the years 1990 and 2000, in which case we have a duration observation, or (b) where the relationship is still ongoing as of 2000, in which case, since the relationship has not ended, we do not have a duration observation except that we know the duration exceeds 10 years. The data in the latter case are referred to as a "right-censored" observation since the end of the relationship (as opposed to the beginning of the relationship, which is referred to as "left censoring") is not observed. Now if one used a regression framework, we can easily deal with the data corresponding to (a) above. However, when the data are right censored, we can either use 10 years as the duration to approximate its true value or we can ignore the observation. As the number of censored observations increases, so will the bias associated with both these approaches. Consequently, the regression model to study durations will not be appropriate when dealing with censored data. Note that left censoring is also an issue with regression models but is less prevalent in marketing data.

The time-varying covariates issues can be thought of as follows. Going back to our illustration on advertising agency relationships, one of the key drivers of the relationship is likely to be the talent pool at the advertising agency. If this talent pool is declining over time, then it will have an impact on the duration of the relationship. Now, in a regression model, the dependent variable will continue to be the duration, but what value of the explanatory variable—size of talent pool—should one use in the analysis? Would it be the pool in 1990, the pool when the relationship terminated, or some average size of the pool over the duration of the relationship? Hence there is no natural way for the regression model to deal with explanatory variables whose values change over the duration. This inability to handle "time-varying covariates" necessitates an alternative approach that can deal with such data.

Having described why standard regression-based models might be unsuitable for studying the effects of explanatory variables or covariates on durations, we now describe the methodology of hazard models. First, why do we refer to these models as "hazard" models? The intuition for this is as follows. What we are interested in at any given point in time is the likelihood that an event will occur given that the event has not occurred thus far. Specifically, in the advertising agency relationship case, one can think of this as studying the probability that the relationship will end at some point in time, given that it has not ended thus far. In other words, we are attempting to assess the risk or hazard of termination of the relationship—hence the term *hazard* models. Equivalently, one can also recast the event of interest in terms of its nonoccurrence until this point in time (i.e., the likelihood of the event occurring after this time point). Since the relationship (in the advertising agency case) has survived until a given point (i.e., the relationship has not yet terminated), these models are also referred to as "survival" models. And since they deal with durations as the dependent variable, they are also called duration models. Following the same logic, it is easy to see why these models are also referred to as failure time models since if a relationship between an ad agency and its client did terminate, then it can be thought of as having "failed" or not having survived.

The notions of hazard and survival described in the previous paragraph must shed some light on why this class of models is able to accommodate right censoring and time-varying covariates. Recall that the issue of right censoring was associated with failure not having occurred by the end of the data collection duration (i.e., the year 2000). What this implies is that the relationship has survived from the start of the relationship in 1990 to the end of data collection in year 2000. So the information contained in this observation will be the likelihood of the relationship having survived at the end of the data collection period. In this way, the methodology can naturally exploit the information contained in right-censored or "survived" observations. In the same way, one can also account for left censoring if it exists in the data on hand.

What, then, about time-varying covariates? To go back to our example, we have annual measures of the size of the talent pool at the ad agency for the years 1990 to 2000, and the pool did decline from one year to the next. Now suppose the relationship ended in the year 1997. Since the explanatory variables only change from year to year, it is straightforward to divide the duration from 1990 to 1997 into subdurations each being a year in length. For the first six subdurations, what we observe is a survival of the relationship given the talent pool in that year or subduration. For the seventh year, we observe a failure (since the relationship terminates) given the talent pool in year 1997. Hence, the overall likelihood of the data will be the product of the likelihoods of the seven subdurations. In this way, the modeling approach accounts for time-varying covariates.

The description in the previous paragraph might prompt the reader to think of a duration model as simply a binary outcome probabilistic model (such as the logit or the probit) where the two outcomes are failure and survival. Such a characterization is accurate except that in a binary outcome model, the intrinsic probability of failure or survival (i.e., the probability without any explanatory variables or covariates) is constant over time. By contrast, in a duration model, this intrinsic probability is allowed to change over time, even if the explanatory variables remain unchanged. Loosely speaking, then, the duration model as described in the paragraph on time-varying covariates can be construed as a binary outcome probabilistic model with time-varying intercepts.

The specification of duration or hazard models (the "hazard function") has three main building blocks: (1) the "baseline" hazard function, (2) a function that reflects the effects of explanatory variables on durations, and (3) an approach to accounting for heterogeneity in the baseline hazard and the covariate function among the various cross-sectional elements whose behavior is being analyzed. Block (3) is mostly relevant only in situations where multiple observations are available for each cross-sectional unit under consideration. Each of the three blocks is discussed in detail in the following.

Baseline Hazard Function

Specifying the baseline hazard is akin to specifying a probability distribution on the duration times in the absence of explanatory variables (see the paragraph on the binary outcome probabilistic model). Since duration times are positive numbers, the standard distributions used for the purpose are the exponential, Weibull, the log-logistic, and the expo-power. Following are the functional forms for these four baseline hazard functions.

$h(t)$ denotes the hazard function, $S(t)$ denotes the survivor function. Using the derivation from Appendix A, one can obtain the relationship between the hazard function $h(t)$ and the survivor function $S(t)$, as $S(t) = e^{-\int_0^t h(u)du}$.

Exponential

$$h(t) = \gamma,$$
$$S(t) = e^{-\gamma t},$$

where $\gamma > 0$. One of the interesting properties of the exponential distribution is that the hazard function corresponding to this distribution is a constant and is the mean parameter of that distribution. This property is referred to as the property of "no duration dependence." In many marketing applications, this property could be unappealing. Consider the case of the advertising agency relationship. What this means is that conditional on the relationship not having terminated until the year 1995, the probability of the relationship terminating is the same as the probability of the relationship terminating conditional on it not having terminated until the year 1999. In other words, the hazard function does not depend on how long the relationship has lasted. In reality, one might think that the hazard function either decreases over time (as the relationship strengthens) or increases over time (as the relationship deteriorates). This would require the distribution to have the property of "duration dependence," such as the following baseline functions.

Weibull

$$h(t) = \gamma\alpha(\gamma t)^{\alpha-1},$$
$$S(t) = e^{-(\gamma t)^{\alpha}},$$

where $\gamma, \alpha > 0$. The Weibull distribution has a hazard that can be increasing in duration if $\alpha > 1$, decreasing over time if $0 < \alpha < 1$, and constant if $\alpha = 1$, which will be the same as the exponential hazard. The shapes of the baseline hazard for exponential and Weibull are shown in Figure 21.1.

Log-Logistic

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

where $\gamma, \alpha > 0$. The log-logistic, in addition, allows for nonmonotonic hazards (i.e., those

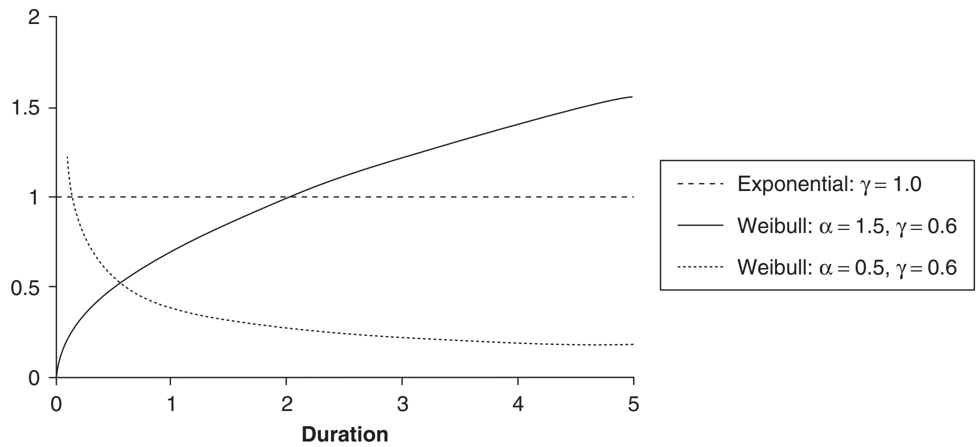

Figure 21.1 Hazard Functions for Exponential and Weibull

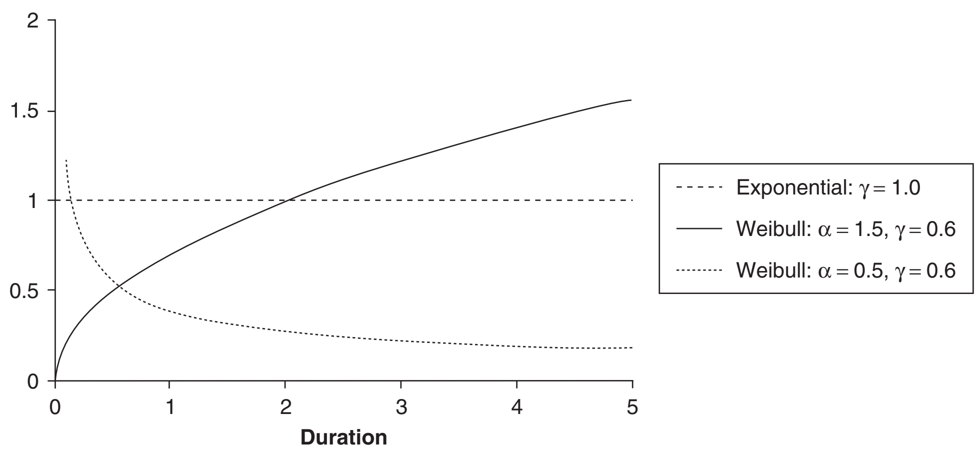

Figure 21.2 Hazard Functions for Log-Logistic With Different Parameter Values

that can increase initially and then decrease). Specifically, for $\alpha > 1$, the hazard first increases with duration, then decreases. Such a specification would be appropriate for data with periodicity, where the likelihood of an event occurring increases over time, but if the event gets delayed beyond a point, it becomes less likely to occur. If $0 < \alpha \leq 1$, the hazard function decreases with duration. The shapes of the functions under different values of α are shown in Figure 21.2.

Expo-Power

$$h(t) = \gamma \alpha t^{\alpha-1} e^{\theta t^{\alpha}},$$
$$S(t) = e^{\frac{\gamma}{\theta}(1-e^{\theta t^{\alpha}})},$$

where γ, $\alpha > 0$. The expo-power is a bit more flexible than the log-logistic and also allows for U-shaped hazard functions. It also nests the Weibull function, as well as the exponential function. The different shapes of expo-power

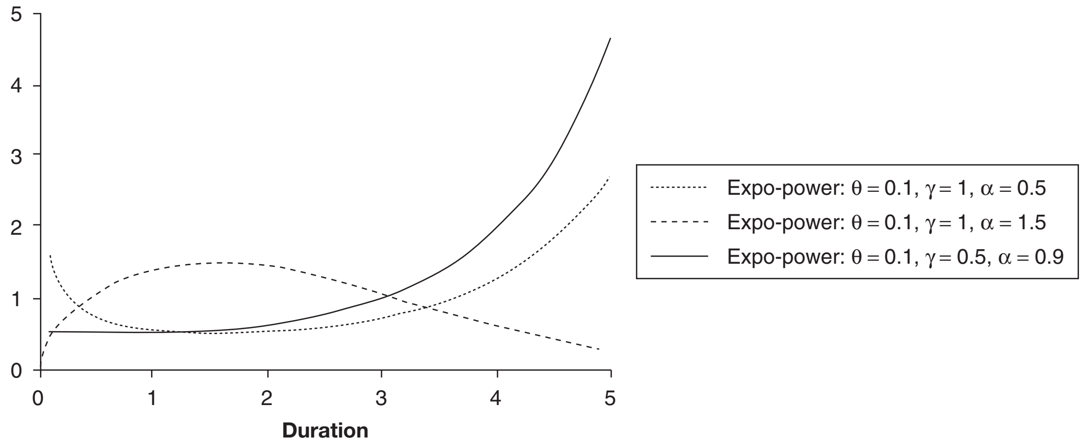

Figure 21.3 Hazard Functions for Expo-Power With Different Parameter Values

with different parameter values are illustrated in Figure 21.3.

Besides these baseline hazard specifications, there are other more flexible forms such as the Box-Cox or the "semiparametric" that involve the estimation of a larger number of parameters. The trade-off for the researcher is the additional flexibility at the cost of increased computational burden. In most practical applications, the parametric forms described above appear to suffice nicely.

Specification of Explanatory Variables' Effects on the Hazard

With the baseline hazard specified, one needs to decide on how to introduce explanatory variables into the specification. There are three popular approaches to doing this, referred to as the proportional hazards model, the additive risks model, and the accelerated failure time model.

1. Proportional hazard model (PHM), proposed by Cox (1972):

$$h_i\,(t, X_t) = h_i(t) \times e^{x_t\beta_i},$$

where $h_i\,(t, X_t)$ stands for household i's hazard function at time t, X_t is a vector of explanatory variables, and β_i is a vector of parameters for household i. In other words, the PHM says that the effect of the function for explanatory variables on the baseline hazard is multiplicative.

2. Additive risks model (ARM), proposed by Aalen (1980):

$$h_i(t, X_t) = h_i(t) + e^{x_t\beta_i}.$$

ARM specifies that the effect of function for explanatory variables on the baseline hazard is additive.

3. Accelerated failure time model (AFTM), proposed by Prentice and Kalbfleisch (1970). AFTM specifies one of the parameters in the baseline hazard function to be a function of explanatory variables. For example, Chintagunta (1998) specifies parameter γ in the log-logistic hazard function to be a linear function of the explanatory variables, that is, $\gamma = \gamma_0 + X_t\gamma_1$.

Essentially, these three specifications differ in the way in which the derivative of the hazard function with respect to the explanatory variable (when it is continuous)—that is, the marginal hazard—depends on the baseline hazard. In the PHM case, the derivative is proportional to the baseline hazard; in the ARM case, it is independent of the baseline hazard; and in the AFTM case, it can be a complicated function of the baseline hazard, typically nonproportional. To understand the intuition underlying these alternative approaches, recall the example of the advertising agency relationships. One might be interested in answering the question, "What happens to the hazard function if I am able to marginally increase the talent pool in a given time period?" The PHM model would predict an effect size proportional to the value of the hazard at that time point. The ARM specification would predict that the corresponding change in the hazard function would not depend on the current duration of the relationship, whereas the nature of impact under the AFTM is not easy to characterize. A priori, there is no theory that can guide the researcher on which approach is best suited in an empirical context. Hence, the choice will have to based on (a) whether the resulting estimates from the model are interpretable as being intuitively plausible and (b) statistical criteria such as fit and ability to predict in a holdout sample.

Unobserved Heterogeneity

Once the baseline hazard and the effects of explanatory variables have been accounted for, the last issue to address is that of accounting for unobserved heterogeneity. In other words, the parameters of the baseline hazard and/or the effects of the explanatory variables need not be the same across all the cross-sectional units in the analysis. By "cross-sectional units," we mean, for example, the agency-client pair in our example of advertising agency relationships; households when we are interested in the interpurchase time behavior of consumers; segments if we are interested in differences in behavior across large-, medium-, and small-sized businesses; and so on. If such heterogeneity exists but is not accounted for in the estimation, it will result in biased estimated effects for the explanatory variables. One example is to study households' purchase behavior, among which households' responses to promotion (or price cut) are believed to differ across households. Two popular approaches can be employed to account for this heterogeneity, as shown in the following.

Discrete Heterogeneity

This approach also has other names, such as latent class, semiparametric, or finite mixture model approach. The underlying assumption here is that there are discrete "segments" in the market, and each segment has a unique set of parameters characterizing the baseline hazard function and/or the effects of explanatory variables.

In the example of households' purchase behavior analysis, one can believe that there are distinct groups of households in their responses to promotion. The number of groups can be decided using statistical criteria. More details of the estimation process are discussed in Appendix B.

Continuous Heterogeneity

The second is the continuous distribution approach, where each unit's parameters are assumed to be drawn from some assumed continuous distribution (such as the normal distribution). The estimation process for this approach is a little more complicated because of its need for simulation. Appendix C shows more details regarding the estimation process of this approach.

While the choice among these approaches is, to some extent, a matter of taste, the important empirical finding is that heterogeneity when present and ignored can severely bias estimated parameters from these models. One can also use statistical approaches to determine which type of heterogeneity assumption is more appropriate for a set of data.

A Numerical Example

Up to now, the three main building blocks for a duration or hazard model have all been discussed; a numerical example from the paper by Seetharaman and Chintagunta (2003) is presented here to illustrate how hazard models are applied empirically. The data are panel data, recording the purchases of 300 panelists in the detergents category over a period of 2 years from June 1991 to June 1993. The empirical distribution of interpurchase time is plotted in Figure 21.4. Interestingly, this plot shows peaks at a multiple of 7 days. This is consistent with the empirical observations that households make shopping trips in weekly intervals (e.g., Dunn, Reader, & Wrigley, 1983; Kahn & Schmittlein, 1989). The modeling process can

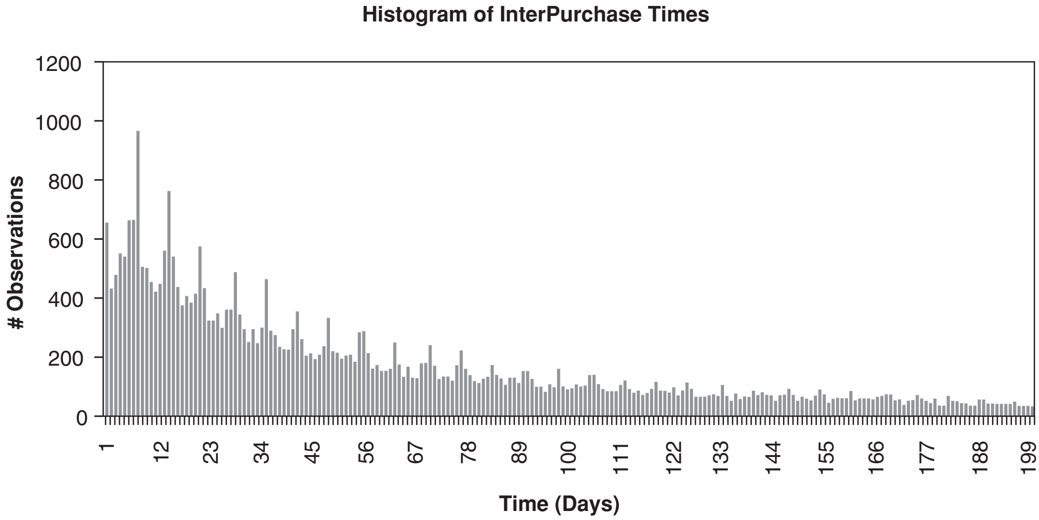

Figure 21.4 Observed Interpurchase Times

be considered as trying to use the three building blocks to separate out their effects on interpurchase times.

The interpurchase time is influenced by marketing variables (price, display, and feature) and a household-specific variable (product inventory). Using the expo-power baseline hazard with discrete time PHM, the parameter estimates for these variables are listed in Table 21.1 for both homogeneous and heterogeneous models. Note that in the homogeneous model, the effect of price is underestimated (closer to zero), relative to price estimates for any segment obtained from the heterogeneous model. This indicates that the model assuming no taste variations among households does not give the average effect on a marketing variable, as one might expect. This can be also verified by the fact that the weighted average of the three estimates from the heterogeneous model (with the probabilities as the weights) is not the same as the parameters from the homogeneous model. The results from the heterogeneous model identify three groups of people. The largest group accounts for 46% among the panelists, who are most sensitive to price and feature. This group could be those "planned shoppers" who have seen the features from their newspapers or online and learned about prices, based on which they decide whether to buy detergents or make a shopping list. The second group accounts for 44% of the panelists, who are least sensitive to price but most sensitive to display. They might be those "unplanned shoppers" who go to the store without a shopping list, but when they see a display and realize that their inventory is low (note that this group shows the highest impact from inventory, compared to the other two groups), they will buy it. The third group accounts for only 10% of the panelists, who are least sensitive to feature or display but only care about price.

Other Types of Hazard Models Encountered in the Marketing Literature

So far, our discussion has focused on a single event that is of interest to marketers, for example, the duration of a relationship between a client firm and an advertising agency. Two other types of durations that have been studied in the marketing literature are as follows. The first situation is one in which, at the end of the duration (i.e., at the time of failure), a number of alternative outcomes are possible. This situation is very commonly encountered with scanner panel data. Consider a household making purchases of instant coffee. When the marketer

Table 21.1 Results Using Expo-Power Baseline Hazard Function

Parameter	*Homogeneous*	*Heterogeneous*		
Price	–0.55 (0.06)	–1.72	–0.61	–1.63
Display	1.41 (0.07)	1.58	1.79	0.85
Feature	1.40 (0.07)	1.94	1.25	0.80
Inventory	–0.02 (0.00)	–0.03	–0.04	0.00
Probability		0.46	0.44	0.10

records that household's purchase (or "failure"), the household could have purchased from among a set of different brands of such coffee—for example, Maxwell House, Folgers, or Taster's Choice. Indeed, the household could have chosen to purchase Maxwell House in the current week because that brand was promoted this week. In the absence of the promotion, the household might have waited an additional week to purchase coffee and even then might have purchased a brand other than Maxwell House. In essence, then, each of the brands is "competing" for the household's purchase. Such a situation in which duration is defined not as the time to purchase instant coffee but rather as the time to purchase the individual coffee brand is referred to as the "competing risks" hazard model. More generally, in the coffee example, if there are J brands the household can choose from, the household at the time of a purchase can be classified as being in one of J "states." Following that purchase, she or he is at risk for again purchasing one of the J brands at the next purchase. In essence there are $J*J$ possible durations of interest to the researcher. Each of these durations (also called "transitions" since they are based on movement from one state to another) can then be modeled exactly as described for the single-duration context above. An important consideration when modeling competing risks is the potentially large number of parameters that need to be estimated when the number of states (or brands in this case), J, gets large. Furthermore, since the transition from state j to state k requires a sufficient number of observations on that transition, we need a large number of parameters across all transitions to be able to estimate all the durations. Nevertheless, the approach allows for an analysis of a household's "transitions" from one state (brand) to another over time. Consequently, a key benefit to this approach is that one can study not just the influence of marketers' actions on the timing of coffee purchases but also the effects of these actions on brand choice behavior.

The other type of hazard model that has appeared in the marketing literature is referred to as the "split" hazard model. Consider an example where we might be interested in characterizing the effects of various factors on time to firms' adoption of a new technology. Given a sample of firms and their durations, we will observe either failures (i.e., adoptions) or survivals (i.e., no adoptions) during the data-gathering period. However, it is possible that some of the survivors are unlikely to ever adopt the new technology, in which case, those firms need to be treated differently from those that are observed not to have made a purchase during the data collection period but could potentially adopt the technology at a later point in time. The split hazard model formalizes this situation by first assigning a probability to each firm of ever adopting the technology. Then, for those that are classified as potential adopters, the duration is characterized very much as described previously (i.e., we can observe either failures or survivors during the data observation period).

Marketing Literature: Some Illustrative Examples

In the marketing literature, there have been two broad classes of applications of duration models. The first set of studies can be thought of being at the "micro" or individual level. The dependent variable of interest here is the time between successive purchases in a frequently purchased product category such as coffee or detergents. Several papers have been published in this area, with a variety of different specifications for the three building blocks described above. The second set of studies tends to be more "macro" in nature and looks at issues pertinent to marketing strategy—what determines the failure of new ventures? When does a new technology "take off," or when does a dominant design emerge in a technology product market? These applications, by and large, have been very successful since researchers have considerable flexibility in specification and estimation of the three key building blocks described previously.

The "classic" study of hazard function or duration models in marketing is that by Jain and Vilcassim (1991). This is the first study in the marketing literature to study households' category purchase timing behavior in the context of the ground coffee category. So the duration of interest was the time to next purchase of the coffee category and the influence of marketing activities such as price and promotion on this duration. It was also the first study to formally decompose the hazard into the baseline hazard, the effects of explanatory variables, and the effects of unobserved heterogeneity as described earlier in this chapter. The authors also tested a number of different parametric specifications for the baseline hazard and concluded that the coffee data, the baseline hazard, had a nonmonotonic pattern. Prima facie, this was a curious finding as one would think that a monotonically increasing hazard is more likely since the probability of making a purchase conditional on not having made a purchase should increase due to the depletion of one's inventory in the pantry. The explanation offered by the authors for the nonmonotonic pattern is based on the regularity of purchases made in this category. In other words, say that a household in this category typically makes a coffee purchase every 4 weeks. Then, it is reasonable (in the absence of any other explanatory variables) to expect the hazard to increase for 4 weeks after a purchase. However, if the regular duration passes for that household, then it becomes less likely to observe the household making a purchase at a later date. Hence, the estimated nonmonotonic nature of the baseline hazard seems reasonable in this case. In terms of accounting for explanatory variables, the authors used the PHM approach. The ability of Jain and Vilcassim to account for unobserved heterogeneity stemmed from the panel nature of the data with repeated observations (also referred to as "multiple spells") on the same household. Indeed, this was also one of the first studies to highlight the importance of accounting for unobserved heterogeneity when investigating household purchase behavior. Similar models published in the marketing include those by Gupta (1991), Gonul and Srinivasan (1993), Helsen and Schmittlein (1993), and Wedel, Kamakura, DeSarbo, and Ter Hofstede (1995). For a review of PHM models published in the marketing literature, the reader is referred to Seetharaman and Chintagunta (2003). For a comparison of some baseline hazard functions, see Grover and Tadikamalla (1997). Wedel et al. (1995) also demonstrate how to account for time-varying covariates with these data by discretizing the duration into subdurations as previously described. A recent application of that approach to online purchasing is in Manchanda, Dube, Goh, and Chintagunta (in press). All the above studies can handle right censoring of the data as well.

While the Jain and Vilcassim (1991) and the studies compare a number of alternative specifications of the baseline hazard model, few studies have compared the different approaches to accounting for explanatory variables in hazard function analyses in marketing. A notable exception is Seetharaman (2004), who compares the performance of the ARM with those of the PHM and AFTM. Using household scanner panel data on three product categories—laundry detergents, paper towels, and toilet tissue—Seetharaman finds that across all product categories, ARM performs the best, followed by PHM and, finally, AFTM. Two criteria are used for this comparison: in-sample fit and prediction in a holdout sample of households. Furthermore, similar to Seetharaman and Chintagunta (2003), he finds that the more flexible functional forms for the baseline hazard fit the data much better than those that impose monotonicity in that hazard, even after adjusting for the larger number of estimated parameters.

In terms of accounting for the effects of unobserved heterogeneity, the early studies (e.g., Jain & Vilcassim, 1991) found that the discrete/latent class approach to accounting for heterogeneity appeared to fit the data better than using specific parametric functional forms for the distribution of this heterogeneity (e.g., the normal distribution). More recently, however, the advent of Bayesian methods (Allenby, Leone, & Jen, 1999; Manchanda et al., in press), and the ability to obtain household-level parameter estimates have reinvigorated proponents of the continuous mixture approach to accounting for heterogeneity.

A few studies in marketing have looked at competing risks models. Again, the first study was that by Vilcassim and Jain (1991). That study investigated the effects of marketing activities on the purchases of saltine crackers. As in the case of coffee category purchases, the finding of nonmonotonic hazards was repeated, albeit at the level of brand-to-brand transitions. The authors again used the PHM framework to account for the effects of the explanatory variables, and they continue to find that it is critical to account for unobserved heterogeneity to obtain the correct pattern for the transitions as well as for the effects of the covariates. An important empirical finding here is the ability to compare switching as well as repeat purchase durations, not only to infer the brands that seem to show the highest loyalty (after controlling for the effects of marketing activities), but also to identify the ability of brands to either draw customers from competing brands or to lose customers to those brands. As identified previously, this methodology gets cumbersome with a large number of brands and also requires a large amount of data to estimate all the transitions. To overcome these problems, Chintagunta (1998) proposes an approach to reduce the computational burden of the competing risks model. The competing risks model has also been applied in the context of a household's personal investment data via a Bayesian framework by Allenby et al. (1999).

The split hazard model has also seen a few applications in marketing. Early studies in this area were by Sinha and Chandrashekaran (1992), who investigated the drivers of bank adoption of ATM machines. More recently, the approach has been applied to study physicians' adoption behavior of new drugs. The popular press (*Wall Street Journal,* July 2005) has recently emphasized the importance of "early adopter" physicians to pharmaceutical companies. By focusing their attention on these physicians, the firm then expects these physicians to act as influencers in getting other physicians to adopt the drug. The question then is, how does one identify the early adopters? In a recent study, Kamakura, Kossar, and Wedel (2004) use time-to-first-prescription data for a large number of physicians across several drugs and estimate a split hazard model on these data. After controlling for the effects of explanatory variables, they are then able to identify physicians who are more inclined to prescribe a drug early. The split hazard formulation is useful in this case to account for the possibility that a physician might never prescribe a particular drug.

Future Directions

Going forward, one can see several directions in which researchers can use and enhance hazard models to better understand marketing phenomena. Very often, marketers have to deal with multiple dependent variables—in some cases, these variables could all be durations, and in other cases, it could be a combination of a duration variable and a dependent variable with different properties (i.e., not a positive valued variable but a discrete variable or a regular continuous variable). A small number of studies in marketing have looked at the issue of multiple dependent variables. For example, when studying the impact of marketing activities on the duration of purchases in two different categories by a household, one needs a bivariate hazard function that is correlated across categories. Chintagunta and Haldar (1998) and, more recently, Park and Fader (2004) have addressed this issue. Researchers have also combined the hazard model with a logit model to study category purchase duration and brand choice behavior of households without having to estimate the competing risks model previously described (see, e.g., Chintagunta & Prasad, 1998). Further researchers have also combined a duration model with a regression model to study the effects of marketing activities on the joint purchase timing and expenditure decisions of households. An illustration of this approach is in Manchanda et al. (in press). Going forward, one expects to see more work in this area.

Substantively, an area where hazard models are just being applied is in measuring the lifetime value of a firm's customers. As the topic of customer relationship management becomes more and more central to the workings of an organization, it is reasonable to expect increased application of hazard/duration models in this area. Indeed, studying customer lifetime value within the context of hazard models poses an interesting methodological challenge. Marketers are interested in influencing two types of durations as they relate to their customer pool. The first is the duration of the relationship—the time elapsed since the inception of the relationship. The second duration involves various events that occur during the span of the relationship such as the different purchases that the customer might make from the firm. Hence, the second duration is embedded or nested within the first duration. One potential framework for addressing this issue is the so-called hazard-in-hazard framework (Lillard, 1993). That framework was developed to account for the timing of births of children within the duration of a couple's marriage. Another potential methodological area of potential investigation (not specific to customer lifetime value) is the area of "excess hazards." The idea is to combine the advantages of PHM and ARM by specifying the hazard as the sum of the PHM (that includes the baseline hazard and covariates) and another baseline hazard. Such a hazard function would be more flexible than either the ARM or PHM. A third area of research could be in the area of "frailty" models. The idea here is that a group of individuals share a common characteristic that influences all their durations. Identifying and estimating this (unobserved) common characteristic can help shed light on the relationship between durations of multiple individuals.

Appendix A: Derivation of the Survivor Function $S(t)$ From the Hazard Function $h(t)$

Based on the definition of hazard, it can be written as $h(t) = \frac{f(t)}{1 - F(t)}$, where $f(t)$ is the probability density function corresponding to the hazard function, and $F(t)$ is the corresponding cumulative distribution function.

It can be also written as $h(t)dt = \frac{dF(t)}{1 - F(t)}$.

Integrating both sides, one can get $\int_0^t h(u)du = -\ln(1 - F(t)) \Rightarrow 1 - F(t) = e^{-\int_0^t h(u)du}$.

Based on the meaning of the survivor function, $S(t) = 1 - F(t)$, that is, $S(t) = e^{-\int_0^t h(u)du}$.

Appendix B: Estimation Method for Model Accounting for Discrete Unobserved Heterogeneity (e.g., There Are G Unobserved Segments in the Market)

The Likelihood Function

Suppose household i belongs to group g and has n_i purchases at time periods $(t_1, t_2, \ldots, t_n)$, with the time variant explanatory variables given by $(X_1, X_2, \ldots, X_n)$. In the analysis, the time variable can be treated as either a continuous or discrete variable. Seetharaman and Chintagunta (2003) provide an extensive review.

In the continuous time case, the likelihood function for household i is $L_{ig} = \prod_{k=1} f(t_k - t_{k-1}, X_k\beta_g)$, where $f(t, X_t\beta) = h(t, X_t\beta) \times S(t, X_k\beta_g)$

$$= h(t, X_k\beta) \times \exp\left(-\int_0^t h(u, X_u\beta)du\right).$$

In the discrete time case, the likelihood function is $L_{ig} = \prod_{k=1}^{T_i} \Pr(t_k, X_k\beta_g)^{\delta_k} \times (1 - \Pr(t_k, X_k\beta_g)^{\delta_k})^{1-\delta_k}$, where T_i is the total number of shopping trips for household i, and δ_k denotes whether this household made a purchase on trip number k (1 if they do, 0 otherwise). $\Pr(t_k, X_k\beta_g)$ is the discrete time hazard for the household at trip k, for household i, who belongs to group g, which can be computed as follows:

a. For PHM, $\Pr(t_k, X_k\beta_g) = 1 - \exp$

$$\left(- \exp(X_k\beta_g) \times \int_{t_{k-1}}^{t_k} h(u)du \right).$$

b. For ARM, $\Pr(t_k, X_k\beta_g) = 1 - \exp$

$$\left(- \exp(X_k\beta_g) + \int_{t_{k-1}}^{t_k} h(u)du \right).$$

The likelihood function, L_i for household i above, is conditional on belonging to segment or group g. The household's unconditional likelihood is given as

$$L_i = \sum_{g=1}^{G} L_{ig} p_g.$$

The weights are the probabilities of anyone falling in each group g, denoted as p_g, and G denotes the total number of groups. Let N denote the total number of individuals in the sample. Then the sample likelihood function is given as follows:

$$L = \prod_{i=1}^{N} L_i.$$

Deciding the Number of Groups *G*

A most commonly used method is based on the Bayesian information criterion (BIC) introduced by Schwarz (1978), which is computed as

$$\text{BIC} = -2 \times LL + K \times \text{In}\ (NOBS),$$

where LL is the log-likelihood value at convergence of the model estimation, K is the number of parameters in the model, and $NOBS$ is the total number of observations in the data. BIC can be seen as a negative function of log-likelihood, with a penalty of the number of parameters as a function of the sample size. Consistent with the concept that a larger log-likelihood value is preferred, a model with a smaller BIC value is preferred.

Starting with $G = 2$, estimate the model using the maximum likelihood estimation method, compute the BIC value, and name it as BIC-2.

Then set $G = 3$, reestimate the model, and compute the BIC value. If this one is larger than BIC-2, then stop and choose $G = 2$ as the optimal number of groups; otherwise, continue to increase the number of groups until the BIC value starts to increase.

Parameter Estimates

Use the parameter estimates from the optimal number of groups, including β_g and p_g, for each group.

Appendix C: Estimation Method for Model Accounting for Continuous Unobserved Heterogeneity

In this case, the βs across different households take different values; β_i is for household i, and the β_is are all from the same distribution (say, normal or multivariate normal). To estimate the parameters of the distributions as well as all the other parameters in the model, one can follow the following four steps:

1. For each household, make NR draws for β_i from the assumed distribution $f(\beta)$, and compute the simulated likelihood function for that household:

$$\tilde{L}_i = \frac{1}{NR} \sum_{r=1}^{NR} L_i(\beta^r),$$

where β_i^r denotes one random draw of β_i.

2. Repeat this for all the households. Note that β_is for all i are drawn from the same distribution, and for the n observations for each household i, β_i is assumed to be constant.

3. Compute the overall likelihood function as a product of all the simulated likelihood functions for each household *i,* that is, $\tilde{L} = \prod_{i=1}^{N} \tilde{L}_i$, where N denotes the total number of households, and $\tilde{L}$ denotes the simulated overall likelihood function.

4. By plugging this function $\tilde{L}$ into a maximum likelihood function procedure, one can get all the parameters, including those for the distributions that can be obtained at convergence.

Note that this method can obtain the distribution only at the population level and not the parameter values for each individual household, the β_is. To obtain the β_is, one can use the hierarchical Bayesian method. For a detailed discussion of that method, please refer to Rossi and Allenby (2003).

REFERENCES

Aalen, O. O. (1980). *A model of nonparametric regression analysis of counting processes.* New York: Springer-Verlag.

Allenby, G. M., Leone, R. P., & Jen, L. (1999). A dynamic model of purchase timing with application to direct marketing. *Journal of the American Statistical Association, 94*(446), 365–374.

Chintagunta, P. K. (1998). Inertia and variety seeking in a model of brand-purchase timing. *Marketing Science, 17*(3), 253–270.

Chintagunta, P. K., & Haldar, S. (1998). Investigating purchase timing behavior in two relates product categories. *Journal of Marketing Research, 35*(1), 43–53.

Chintagunta, P. K., & Prasad, A. R. (1998). An empirical investigation of the "dynamic McFadden" model of purchase timing and brand choice: Implications for market structure. *Journal of Business and Economic Statistics, 16*(1), 2–12.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society B, 34,* 187–220.

Dunn, R., Reader, S., & Wrigley, N. (1983). An investigation of the assumptions of the NBD model applied to purchasing at individual stores. *Applied Statistics, 32*(3), 249–259.

Gonul, F., & Srinivasan, K. (1993). Consumer purchase behavior in a frequently bought product category. *Journal of the American Statistical Association, 88*(424), 1219–1227.

Grover, R., & Tadikamalla, P. (1997). A comparison of some stochastic models of interpurchase intervals and of aggregate purchase incidence. *American Journal of Mathematical and Management Sciences, 17,* 299–328.

Gupta, S. (1991). Stochastic models of interpurchase time with time-dependent. *Journal of Marketing Research, 28,* 1–15.

Helsen, K., & Schmittlein, D. C. (1993). Analyzing duration times in marketing: Evidence for the effectiveness of hazard rate models. *Marketing Science, 12*(4), 395–414.

Jain, D. C., & Vilcassim, N. J. (1991). Investigating household purchase timing decisions: A conditional hazard function approach. *Marketing Science, 10*(1), 1–23.

Kahn, B. E., & Schmittlein, D. C. (1989). Shopping trip behavior: An empirical investigation. *Marketing Letters, 1*(1), 55–69.

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

Lillard, L. (1993). Simultaneous equations for hazards: Marriage duration and fertility timing. *Journal of Econometrics, 56,* 189–217.

Manchanda, P., Dube, J.-P., Goh, K., & Chintagunta, P. (in press). The effect of banner advertising on internet purchasing. *Journal of Marketing Research.*

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

Prentice, R. L., & Kalbfleisch, J. D. (1970). Hazard rate models with covariates. *Biometrics, 35*(1), 25–39.

Rossi, P., & Allenby, G. M. (2003). Bayesian statistics and marketing. *Marketing Science, 22,* 304–328.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461–464.

Seetharaman, P. B. (2004). Modeling multiple sources of state dependence in random utility models: A distributed lag approach. *Marketing Science, 23*(2), 263–271.

Seetharaman, P. B., & Chintagunta, P. K. (2003). The proportional hazard model for purchase timing:

A comparison of alternative specifications. *Journal of Business and Economic Statistics, 21*(3), 368–382.

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

Vilcassim, N. J., & Jain, D. C. (1991). Modeling purchase-timing and brand-switching behavior incorporating explanatory variables and unobserved heterogeneity. *Journal of Marketing Research, 28,* 29–41.

Wedel, M., Kamakura, W. A., DeSarbo, W. S., & Ter Hofstede, F. (1995). Implications for asymmetry, nonproportionality, and heterogeneity in brand switching from piece-wise exponential mixture hazard models. *Journal of Marketing Research, 32,* 457–462.