Modeling Purchases as Repeated Events*

Govert E. Bijwaard†
SEOR
Econometric Institute
Erasmus University Rotterdam

Philip Hans Franses
Econometric Institute
Department of Marketing and Organisation
Erasmus University Rotterdam

Richard Paap
Econometric Institute
Erasmus University Rotterdam

Econometric Institute Report EI 2003-45

Abstract

We put forward a statistical model for interpurchase times that takes into account all the current and past information available for all purchases as time continues to run along the calendar timescale. It delivers forecasts for the number of purchases in the next period and for the timing of the first and consecutive purchases. Purchase occasions are modeled in terms of a counting process, which counts the recurrent purchases for each household as they evolve over time. We show that formulating the problem as a counting process has many advantages, both theoretically and empirically. We illustrate our model for yogurt purchases and we highlight its useful managerial implications.

Key words: Purchase timing; repeated events; counting process; Cox model; unobserved heterogeneity.

*We thank Dennis Fok for data support, and Pradeep Chintagunta for helpful discussions.

†Corresponding author: Erasmus University Rotterdam, Econometric Institute, P.O. Box 1738, NL 3000 DR Rotterdam, The Netherlands; Phone: (+31) 10 40 82819; Fax: (+31) 10 40 89162; E-mail: `bijwaard@few.eur.nl`

1 Introduction

Household panel scanner data contain information on household-specific purchase behavior, including brand choice, unit purchases and interpurchase times. This last variable is important for store managers, as it allows for an analysis of the timing of purchase behavior. That is, such data can be informative for dynamics of purchases and it allows for an examination of the covariates that increase or decrease the time between purchases. From a managerial point of view, one might use such data to decide on the timing between promotional actions. Also, these data can be relevant for active stock management, where knowledge of depletion rates might allow a store manager to cut back costs.

The currently applied models for interpurchase timing describe the time until the next purchase, see Gupta (1991), Jain and Vilcassim (1991), Vilcassim and Jain (1991), Helsen and Schmittlein (1993) and Chintagunta and Prasad (1998) among others for an application, and see Seetharaman and Chintagunta (2003) for a recent overview. Upon assuming a hazard function, the interpurchase time is correlated with various explanatory variables which are sometimes allowed to change during the duration process, see Gupta (1991). An important property of these models is that, after the purchase has been made, time is reset to zero and a new duration spell starts, independent of the previous duration. In fact, each interpurchase time of a household is modeled separately, using the same type of hazard function.

Considering each purchase timing as an independent random variable has several limitations, in particular if one wants to use the model for managerial decisions. First of all, we neglect valuable information about the purchase history of a household's decisions. Not only is time reset to zero, also the integrated hazard function, which reveals information about past purchase decisions, is not taken into account. This function may, for example, reflect the sensitivity of households to changes in the marketing-mix variables observed during and before the previous interpurchase time. Of course, the information from previous purchase decisions may be very valuable for predicting future interpurchase times and for analyzing the effects of promotional activities on future purchase behavior, and hence, sometimes one does not want to loose that information. In fact, resetting time to zero, as is common in all currently available models, removes much of the typical behavior of the interpurchase process in a similar way as first differencing does for times series data. As a consequence, this reset approach imposes limitations for the use of time-varying covariates and duration dependence in the models.

Second, as the current models only describe the time until the next purchase, the models can only be used to forecast one purchase ahead. Hence we can use the models to compute the probability that the next purchase is made in the next week or in the next two weeks, see for example Gupta (1991). Unfortunately, the models cannot be used to predict whether a household will make two or more purchases in, say, the next two weeks, and this amounts to a serious limitation of the current models. Indeed, managers would like to know the effect of promotional activities on the number of purchases in the current week but also on that in subsequent weeks.

In this paper we therefore introduce a statistical model for repeated duration data, which solves the problems described above. In contrast to the standard approach, where time is reset after a purchase, it takes into account all information concerning all past purchase occasions. The time is not reset once a purchase is made, but it continues to run along the calendar timescale. The decision to purchase at one point in time now depends on the whole path of the purchase history starting at the beginning of the observation window. The model also delivers forecasts for the number of purchases in the next period as well as for the timing of the first and consecutive purchases.

The intuition behind our model is as follows. Purchase occasions are modeled in terms of a counting process, which counts the purchases for each household as they evolve over time. We show that formulating the problem as a counting process has many advantages. The counting process formulation concerns the situation where a household is at risk of purchasing an item by defining an at-risk-indicator. If, for example, a household is only observed after its first purchase, the at-risk-indicator is zero until this first purchase. The incorporation of time-varying covariates is straightforward in the counting process formulation. As we will argue, the joint use of different time scales in the analysis, like (i) the time since the observations started, (ii) the time since the previous purchase and (iii) the calendar time, is easy to implement. It is therefore also easy to model seasonal effects in purchase timing, if necessary. The counting process formulation visualizes the close connection between models for counts, like a Poisson model or Negative Binomial model, and the familiar duration models. Purchase quantities can be included in the model in a natural way. Basically, we advocate the use of the method described in Andersen and Gill (1982), where in our paper we extend it to include unobserved heterogeneity.

The outline of our paper is as follows. In Section 2, we describe the counting process view on repeated events. We discuss all its features and their implications for modeling (re)purchase timing. In Section 3, we describe this model specific for interpurchase times and we discuss in detail how this can be implemented and which type of forecasts it can deliver. We discuss parameter estimation, where we specifically focus on the case of unobserved heterogeneity. When relevant, we compare our model with close competitors. In Section 4, we illustrate the model for a household scanner panel data set on purchases of yogurt. Graphs are used to show how the model outcomes can be used for managerially relevant purposes. In Section 5, we conclude and we discuss avenues for further research.

2 Introducing counting processes

To introduce the modeling of recurrent event data, we first discuss some notation. Suppose that we have a sample of $i = 1, \dots, N$ households. For each household we observe recurrent purchases within a certain product category, denoted by $j = 1, \dots, k_i$, where k_i denotes the number of purchases of household i in the sample. Let T_{ij} be the occurrence time of the jth purchase of household i measured from the start of the observation period. We allow for delayed entry in the study and denote the occurrence time of entry by S_i. This is necessary as we only observe a household after it has purchased the product for the first time. Hence, for each household we observe a sequence of purchase occasions: $0 \leq S_i < T_{i1} < \dots < T_{ik_i}$. These k_i purchase times are uncensored, which is in contrast with the $(k_i + 1)$th purchase time, starting at T_{ik_i}, which is censored. Hence, we only know that the duration is larger than $T_e - T_{ik_i}$, where T_e denotes the occurrence time of the end of the observation period.

2.1 Risk intervals and risk sets

To describe recurrent event data we need to consider two important concepts, that is, the *risk interval* and the *risk set*. The risk interval corresponds to the time interval where a household is at risk of purchasing. The risk set is the collection of households which are at risk at a certain point in time.

The definition of the risk interval depends on the timescale which is used to describe the data. There are several ways to deal with time, where we follow Kelly and Lim (2000) who distinguish between *total time*, *gap time*, and *calendar time*, see also Duchateau et al. (2003). Figure 1 displays the three situations for the purchase history of two households. The purchase history is denoted in panel (a) of the figure.

In the gap-time representation, shown in panel (c) of Figure 1, time at risk starts at 0 after a purchase (or entry to the study in case of the first purchase) and ends at the time of the next purchase. Hence, time is reset to zero after each purchase. Note that this is the starting point for the conventional interpurchase times models in marketing, see for example, Gupta (1991), Jain and Vilcassim (1991) and Helsen and Schmittlein (1993).

Another view is that the time at risk for a particular purchase in the total time representation starts at 0 when the household enters the study and it lasts until the particular purchase, see panel (d) in Figure 1. This representation has less intuitive appeal when the households enter the study at different times because then we have a different timescale for each household. This specification is therefore not suited for interpurchase times applications.

Finally, in the calendar time formulation, the length of the time at risk is the same as in the gap-time representation. The only difference is however that the starting time of the at-risk period is not reset to zero after a purchase but it is put equal to the actual time since the beginning of the observation period, see panel (b) of Figure 1.

To analyze purchase timing, the evolution of the whole purchase history and the impact of current and past marketing-mix variables on purchase timing decisions is important.

As can be seen from Figure 1, a model based on the calendar time representation seems to be the best approach. In contrast to the gap-time representation, a model based on calendar time does not neglect valuable information about the purchase and marketing-mix history of the households nor does it give problems with an interpretation of the purchase durations when households enter the study at different times. Another advantage of the calendar time representation is that calendar time effects, like seasonal effects or day-of-the-week effects, are easy to implement, simply because they follow the same time path. For the other timescales, incorporating seasonal effects is difficult and it will also be non-trivial to interpret these effects.

To discuss the risk interval and the risk set for the calendar time representation in detail, we again consider the observed purchase history for two hypothetical households in Figure 1. Household 1 is observed from the start and makes a purchase in, say, week 5, 14, and 24 and the observation period ends in 29 weeks. Hence, we have to deal with right censoring. Household 2 enters the study after 4 weeks and makes a purchase after 16 and 18 weeks and its duration time is also censored at week 29.

In household scanner panel data we usually observe the purchase behavior of households within a predefined observation period. If the purchase history of all households is observed in this period, the risk interval for each household is equal to the observation window. This is the case for the first household in Figure 1. When households are only observed after the first purchase moment (like the second household), the risk interval begins right after this first purchase moment. The purchase data of these households are truncated from the left. If a household is removed from the sample before the end of the observation period, for example because it left the panel, the risk interval ends at the time the household leaves the sample. In situation (b), we can even consider discontinuous risk intervals in which households can be at-risk or off-risk at alternating intervals. For recurrent events, like repurchasing a product, a household is not considered to be at risk for the jth purchase until it has purchased the product $(j-1)$ times, see panel (b) of Figure 1. For example, the second household purchasing a product for the first time at week 4 (entry date) and repurchasing at week 16 and 18, is at risk for the first purchase from week 4 until week 12, and for the second purchase from week 12 until week 14.

The risk intervals determine the risk sets for each purchase. The risk set contains all households that are at risk for a particular purchase. For standard survival data (single event data), the risk set at a particular time typically consists of all households that have entered the study and that are still observed at that time. For recurrent purchases, however, we can distinguish between restricted and unrestricted risk sets. In case of unrestricted risk sets we consider each purchase as a similar event. That is, at each point in time the risk set for a repurchase at that time consists of all households currently observed. It does not matter how many purchases the households already made at this point in time. If the number of the purchases already made is considered to be important, a restricted risk set is used. Contributions to the jth risk set are restricted to include only households who already made $j-1$ purchases.

To illustrate both concepts, consider the time period between the second and third purchase of the first household in panel (a) of Figure 1. If we consider unrestricted risk sets, both households are in the risk set at each point in time in the interval. However, if we consider a restricted risk set, the second household only enters the risk set of the second purchase after it has made its second purchase. The second household does not belong the second purchase risk set in the period between the second purchase of the first household and the second purchase of the second household.

For standard household scanner panel data where households frequently purchase within a product category, the unrestricted risk sets are relevant. Restricted risk sets may become relevant if, for example, a household receives a bonus after purchasing three items within the category within a certain period of time, or if one wants to analyze the penetration of a new product in the market.

2.2 Features of counting processes

Using the risk intervals and risk sets, we can formulate a stochastic counting process which describes the number of repurchase occurrences. Although counting processes may look complicated at first sight, we show that many aspects of repurchase data can easily be understood if one uses counting processes. Andersen, Borgan, Gill, and Keiding (1993) provide an excellent survey of counting process theory. A less technical survey is given in Klein and Moeschberger (1997) and Therneau and Grambsch (2000).

To start the discussion, we first introduce some notation. A *counting* process $N_i(t)$ is a stochastic process which describes the number purchases of household i in the interval $[0, t]$ as time proceeds. The process has only jumps of size $+1$. This implies that for each household only one purchase can be made at any point in time. The counting process formulation is in close connection with models for count data, like a Poisson model or a Negative Binomial model, and with duration models.

The counting process is governed by its random *intensity* process $\lambda_i(t)$. If we consider a small interval $(t - dt, t]$ of length dt, then $\lambda_i(t)dt$ is the conditional probability that $N_i(t)$ jumps in that interval given all that has happened until just before t. Let $dN_i(t)$ denote the increment of $N_i(t)$ in the small interval and let $\mathcal{H}_{it}$ denote the information set of household i up to, but not including, t. This *history* process includes a complete specification of the path of the counting process on $[0, t)$ and it includes all other events implicitly or explicitly included in the model which have happened before time t. Hence, the history for household i also includes the occasions when this household was at risk. Let $Y_i(t)$ be an indicator function which is 1 if household i is at risk at time t, and 0 elsewhere. Note that we have $\mathcal{H}_{is} \subseteq \mathcal{H}_{it}$ for all $s \leq t$, which indicates that the information set increases over time. We can write the conditional probability that household i makes a purchase at time t given its history as

$$\Pr[dN_i(t) = 1|\mathcal{H}_{it}] = Y_i(t)\lambda_i(t)dt. \tag{1}$$

Note that $\mathcal{H}_{it}$ contains past information about the value of $Y_i(t)$. Hence, the history of being at risk in the past is completely captured in $\mathcal{H}_{it}$. In the calendar time representation a household is only "at risk" of purchasing the product after entry into the study, which occurs at its first purchase S_i, until the end of the study, T_e. A household is only "at risk" of purchasing the product for the jth time after having purchased it for already $(j-1)$ times. This implies that the at-risk indicator for the jth purchase is $Y_{ij}(t) = I(T_{ij-1} < t \leq T_{ij})$, with $T_{i0} = S_i$.

The individual counting processes can be aggregated to $N(t) = \sum_{i=1}^{n} N_i(t)$, which represents the total number of purchases in $(0, t]$. Aggregation of the risk intervals leads to the risk sets at each point in time on the calendar timescale. The risk set at a particular time is the number of households at risk of repurchasing at that given time. With a restricted risk set only the risk intervals of the jth purchase contribute to the risk set of the jth purchase at a given time. The size of the risk set at time t is given by $Y_j(t) = \sum_{i=1}^{n} Y_{ij}(t)$. Thus, at each time the risk set for the first purchase differs from the risk set of the second (and later) purchase. Although this might be a plausible assumption for the analysis of the market penetration of a new product, for the analysis of the repurchase of a frequently purchased product it does not. For our analysis we therefore assume that the risk intervals for all purchases contribute to the risk set of the jth purchase, hence the risk set is unrestricted. The size of the risk set at time t is then $Y(t) = \sum_{i=1}^{n} Y_i(t)$ with $Y_i(t) = \sum_{j=1}^{k_i+1} Y_{ij}(t)$. This implies that, in typical repurchase data with delayed entry (after the first purchase), the number of households at risk at time t is equal to the number of households who have purchased the item at least once up to t.

Another important process is the *cumulated intensity*

$$\Lambda_i(t) = \int_0^t \lambda_i(s)\, ds. \tag{2}$$

This process measures the expected number of repurchases of household i at time t given the observed history up to, but not including, t. Thus,

$$\mathrm{E}[N_i(t)|\mathcal{H}_{it}] = \Lambda_i(t). \tag{3}$$

Using (3) we can derive the expected number of purchases within each given interval, say from t_1 until t_2 (with $t_2 > t_1$), which is equal to $\Lambda_i(t_2) - \Lambda_i(t_1)$.

In general, the intensity function will depend on the current and past value of the marketing-mix variables. If a manager knows the (planned) marketing-mix schedule for, say, the coming month, (s)he can predict the impact of these marketing-mix variables on the expected purchases (in number and timing) in the next month. Note that this is only possible due to the fact that we have chosen for a calendar time model. Indeed, in gap-time models one cannot predict whether a household will make two or more purchases in the next month. Hence, forecasting the impact of marketing-mix variables on the purchases in a fixed period ahead is difficult, if not impossible, with gap-time based models, but not using our formulation.

3 Modeling repeated interpurchase times

In the previous section we have introduced counting processes. We have not yet considered explicitly how covariates, like marketing-mix variables and household characteristics, can be included in counting process modeling. For ordered repeated events, that is, multiple events of the same type, several suggestions have been made to analyze the effect of observed covariates on the reoccurrence intensity $\lambda_i(t)$. A major issue for modeling repeated events is the treatment of possible correlation between the events of one household, that is, the correlation between the interpurchase times.

The three most common approaches are the independent increment model of Andersen and Gill (1982), the marginal model of Wei, Lin, and Weissfeld (1989), and the conditional model of Prentice, Williams, and Peterson (1981). All three approaches are based on a "marginal" regression model in the sense that the regression parameters are determined ignoring the correlation between the events of the same household. Afterwards the standard errors of the estimates are corrected for this assumption.

The three approaches differ considerably in their creation of the risk sets. For the marginal and conditional models, each occurrence of the event (a yogurt purchase in our application) is modeled as a separate event, while the Andersen and Gill (1982) [AG] model assumes that all purchase events are identical. With our application in mind, modeling each purchase separately is not a sensible choice, because it is not plausible that for a frequently purchased product as yogurt the purchase intensity changes with the number of purchases. Therefore, we only consider the AG model and its extension with unobserved heterogeneity. In this extension the correlation between the purchases of the same household is captured through a household-specific random effect.

3.1 The Andersen-Gill Counting Process Model

The method of Andersen and Gill (1982) is the closest in spirit to a Poisson regression model (Lawless 1987; Winkelmann 1995), and it can in fact be accurately approximated with Poisson regression methods. In particular, in the AG model the distribution of the cumulated intensity $\Lambda_i(t)$ uniquely determines the distribution of the number of repurchases, the counts of reoccurrence. This distribution is a (non)-homogeneous Poisson distribution.

Let $x_i(t)$ be a vector of possibly time-varying covariates of household i at time t and let $\bar{X}_i(t) = \{x(s) : s \leq t\}$ denote the complete path of the covariate vector up to time t. For a time-constant covariate we have that $\bar{X}_i(t) = X_i$. The $\bar{X}_i(t)$ set may contain, for example, the whole path of the marketing-mix variables until t, but it may also contain household characteristics. In the AG model, the intensity process for the ith household given these covariates is

$$Y_i(t)\lambda_0(t)\exp(\beta' X_i(t)), \tag{4}$$

where $\lambda_0(t)$ is the *baseline intensity* and β is a parameter vector. The covariates have a multiplicative effect on the intensity through a log-linear regression function $\exp(\beta' X_i(t))$. This model is the natural extension to recurrent events of the familiar proportional hazard model (also called Cox regression model) for survival data. In such a set-up, $\lambda_0(t)$ is the same for all households and it is called the *baseline hazard*. The difference between the proportional hazard model and the AG model lies in the definition of the risk indicator $Y_i(t)$. In the standard proportional hazard model, the household ceases to be at risk after the purchase of the product and hence $Y_i(t)$ becomes zero. In the AG model for recurrent events $Y_i(t)$ remains one as long as household i has not left the study.

The AG model imposes that for each household the subsequent purchases are mutually independent. This is due to the fact that each household's counting process has independent increments. Such processes are typically modeled as time-varying Poisson processes. This assumption can be relaxed by introducing time-varying covariates in the model, such as the time since the previous purchase, which may capture the dependence structure among the successive purchase times.

As Seetharaman and Chintagunta (2003) point out, there is no prescription for which baseline intensity function is the most appropriate to characterize repurchase times. In the past a variety of baseline intensity formulations have been used, such as Weibull, Erlang-2 and exponential power. In Cox models the baseline intensity is treated as a nuisance parameter and estimated non-parametrically, see Cox (1972). The baseline hazard can, therefore, exhibit any shape. Helsen and Schmittlein (1993) discussed this Cox partial likelihood estimation procedure in the context of duration times in marketing. Note that the baseline intensity in models on the calendar timescale have a different interpretation than it has for the gap-time models. On the calendar timescale, the baseline intensity represents seasonal and day-of-the week effects. Hence, if on a particular day of the week much more yogurt is purchased than on other days, this will appear in the intensity process as peaks of multiples of 7 days. In the gap-time models, the baseline intensity represents the duration dependence from the previous purchase. Fortunately, it is easy to have duration dependence from previous purchases in the AG model. For example, Weibull time dependence is incorporated by adding the logarithm of the duration since the previous purchase to the explanatory variables. The parameter of this (time-varying) covariate is equal to the standard Weibull parameter minus one. In a similar way, we can derive functions of time since the previous purchase to reflect other types of duration dependencies.

Cox (1972) suggests that inference on the regression parameters in the proportional hazard model can be based on a partial likelihood function. In this approach one considers the conditional probability that a purchase is made by household i at time t given that each household at risk could have made a purchase at that time, that is,

$$\frac{dN_i(t)\lambda_0(t)\exp(\beta' X_i(t))}{\sum_{l=1}^{n} Y_l(t)\lambda_0(t)\exp(\beta' X_l(t))}. \tag{5}$$

Due to the multiplicative nature of the AG model, the baseline intensity drops out off this probability. If we neglect censoring, the partial likelihood in the case of repurchase times is the product of the conditional probabilities of all purchase times T_{ij} of the households[1]

$$\ell(\beta) = \prod_{i=1}^{n} \prod_{j=1}^{k_i} \frac{\exp\bigl(\beta' X_i(T_{ij})\bigr)}{\sum_{l=1}^{n} Y_l(T_{ij}) \exp\bigl(\beta' X_l(T_{ij})\bigr)} \tag{6}$$

The main difference with the conventional partial likelihood is that we have the product of k_i repurchase times, instead of just one (survival) time. The parameter β can be estimated by maximizing the partial likelihood function with respect to β using numerical methods. The covariance matrix of the estimator is the inverse of the information matrix of the partial likelihood function. Parameter estimation of this model is available in SAS and STATA, see Appendix A.

An estimate of the baseline intensity $\lambda_0(t)$ is very similar to the Breslow (1972) estimate of the baseline hazard for survival time at each time t_k a purchase takes place, is equal to

$$\hat{\lambda}_0(t_k) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{k_i} dN_{ij}(t_k)}{\sum_{i=1}^{n} Y_i(t_k) \exp\bigl(\hat{\beta}' X_i(t_k)\bigr)} \tag{7}$$

The cumulated intensity at time t is estimated as the sum of the baseline intensities up to time t,

$$\hat{\Lambda}_0(t) = \sum_{t_k \leq t} \hat{\lambda}_0(t_k) \tag{8}$$

If there is interdependence of the recurrent purchases due to omitted covariates or household-specific effects, the parameter estimates may be biased and/or the estimated covariance matrix provides invalid standard errors. To correct for this, one may use a robust covariance matrix estimate, see Lin and Wei (1989). Another approach is to explicitly model the household-specific effects using unobserved heterogeneity, which we will do in the next subsection. One adds to the intensity process a household-specific latent variable, which has a specified parametric distribution. Models with such unobserved heterogeneity deal with the correlation among repurchase times by using a random effect term. In Section 3.2, we consider the AG model with Gamma distributed unobserved heterogeneity to account for household-specific effects.

[1]If households purchase at the same time, the partial likelihood functions have to be adjusted using the method close to the Breslow (1974) or Efron (1977) adjustment for ties for survival data. For notational clarity we assume that the data do not contain such ties.

The covariance correction approach uses the fact that, in the presence of misspecification, the standard AG model estimator for the regression parameters converges to a well-defined value that can be interpreted meaningfully. To account for the interdependence from household-specific effects, the covariance matrix of the estimator is adjusted. The adjusted covariance matrix of the estimator is based on the assumption that observations are independent across households but not necessarily within households. The resulting robust covariance matrix estimator is given by

$$V_R = V^{-1}(S'S)V^{-1}, \tag{9}$$

where V is the inverse of the information matrix belonging to (6) and S is the matrix of first derivatives of the logarithm of the likelihood contribution per individual i.

3.2 Unobserved heterogeneity

Interdependence of consecutive purchase times can also be induced by unobserved characteristics of the households, like being a heavy-user or not. If we do not account for these possible missing variables, the parameter estimator may be biased, see Lancaster (1979). It warrants inclusion of a household-specific effect in the model. Such effects are often referred to as unobserved heterogeneity. In Cox survival models this kind of model is called the mixed proportional hazard model.[2] The intensity process of household i at time t is now given by

$$Y_i(t)v_i\lambda_0(t)\exp(\beta' X_i(t)), \tag{10}$$

where the v_i are i.i.d. random variables with distribution function $G(v)$, see Oakes (1992) among others. Because the intensities are non-negative, the distribution of the unobserved heterogeneity is usually chosen from the class of non-negative distributions. In practice, one usually opts for Gamma, Log-normal or Stable distributions, with Gamma being the most popular one. Conditional on the chosen parametric distribution, the interpurchase times are assumed to be independent.

[2] In the biomedical literature these kind of models are called frailty models.

Here, we assume that the unobserved heterogeneity has a Gamma distribution with mean 1 and variance θ, that is, the density function is given by

$$G(v) = \frac{v^{(1/\theta-1)}\exp(-v/\theta)}{\Gamma\big(1/\theta\big)\theta^{1/\theta}}. \tag{11}$$

Hence, large values of θ reflect a greater degree of heterogeneity among households and a stronger association within household purchases. The log-likelihood is given by

$$L(\beta,\lambda_0,\theta) = \sum_{i=1}^{n}\Bigg\{k_i\ln(\theta) - \ln\Gamma\Big(\frac{1}{\theta}\Big) + \ln\Gamma\Big(\frac{1}{\theta}+k_i\Big)$$
$$-\Big(\frac{1}{\theta}+k_i\Big)\ln\Big[1+\theta\int_{S_i}^{T_e}\lambda_0(s)\exp\big(\beta' X_i(s)\big)\,ds\Big] + \sum_{j=1}^{k_i}\big[\beta' X_i(T_{ij}) + \ln\big(\lambda_0(T_{ij})\big)\big]\Bigg\}, \tag{12}$$

where S_i is the entry time of household i and T_e is the end of the observation period (which the same for all households). Note that we assume that we observe all purchase times except for the last one, which occurs after the observation period has ended. Recall that k_i has been defined as the number of observed purchases by household i.

3.3 The estimation procedure

If we assume that the baseline intensity $\lambda_0(t)$ has some parametric form, we can directly maximize the likelihood function (12) to obtain parameter estimates. Estimates of the variance-covariance matrix are obtained by evaluating the inverse of the information matrix in the parameter estimates.

If we do not assume a parametric form for the baseline intensity, semi-parametric parameter estimates can be obtained by using an EM algorithm, see Dempster, Laird, and Rubin (1977). To estimate the model parameter using EM, we modify the procedures in Klein (1992) to allow for time-varying covariates and delayed entry.[3] First, note that if we could observe the v_i's, up to the integrating constant the log-likelihood function is given by

$$L(\beta,\lambda_0,\theta|\text{data}, v_1,\ldots,v_n) = L_1(\theta) + L_2(\beta,\lambda_0),$$

[3] Nielsen et al. (1992) present an alternative EM estimation scheme. Another alternative is to use a penalized likelihood method, see Therneau and Grambsch (2000) and Rondeau, Commenges, and Jolly (2003).

where

$$L_1(\theta) = -n\Big[\frac{\ln(\theta)}{\theta} + \ln\Gamma\Big(\frac{1}{\theta}\Big)\Big] + \sum_{i=1}^{n}\Big\{\Big(\frac{1}{\theta} + k_i - 1\Big)\ln(v_i) - \frac{v_i}{\theta}\Big\}, \tag{13}$$

and

$$L_2(\beta, \lambda_0) = \sum_{i=1}^{n}\sum_{j=1}^{k_i}\Big[\beta' X_i(T_{ij}) + \ln\big(\lambda_0(T_{ij})\big)\Big] - \sum_{i=1}^{n} v_i \int_{S_i}^{T_e} \lambda_0(s)\exp\big(\beta' X_i(s)\big)\,ds. \tag{14}$$

This function is called the complete data log-likelihood function. Klein (1992) provides the outline of the EM algorithm. In the Estimation-step (E-step), one computes the expected value of the complete data log-likelihood function with respect to v, given the current estimates of the parameters and the observed data. In the Maximization-step (M-step) the expected complete data log-likelihood function obtained from the E-step is maximized with respect to the unknown the parameters. The algorithm iterates between these two steps until convergence. To initialize the EM algorithm, we make an initial guess of the values of the parameters. For example, one may put θ equal to zero and use the partial likelihood procedure for the standard AG model based on (6) and (7) to obtains starting values for β and λ_0.

To apply the E-step of the EM-algorithm, we use the fact that, conditional on the observed data, the v_i's are independent Gamma distributed random variables with shape parameters $A_i = 1/\theta + k_i$ and scale parameters

$$C_i = 1/\theta + \int_{S_i}^{T_e} \lambda_0(s)\exp\big(\beta' X_i(s)\big)\,ds.$$

This results in the expected log-likelihood functions L_1^E and L_2^E given by

$$L_1^E(\theta) = -n\Big[\frac{\ln(\theta)}{\theta} + \ln\Gamma(\frac{1}{\theta})\Big] + \sum_{i=1}^{n}\Big\{\Big(\frac{1}{\theta} + k_i - 1\Big)[\psi(A_i) - \ln(C_i)] - \frac{A_i}{\theta C_i}\Big\}, \tag{15}$$

where $\psi(\cdot)$ denotes the Digamma function, that is, the derivative of the Gamma function $\Gamma(\cdot)$, and

$$L_2^E(\beta, \lambda_0) = \sum_{i=1}^{n}\sum_{j=1}^{k_i}\Big[\beta' X_i(T_{ij}) + \ln\big(\lambda_0(T_{ij})\big)\Big] - \sum_{i=1}^{n} \frac{A_i}{C_i} \int_{S_i}^{T_e} \lambda_0(s)\exp\big(\beta' X_i(s)\big)\,ds \tag{16}$$

The M-step of the EM algorithm requires maximization of (15) and (16) with respect to the unknown parameter θ and β. The latter is just the complete log-likelihood function of a Cox model with a additional household-specific covariate, $\ln(A_i/C_i)$, with a known coefficient of one. We can therefore apply a Cox partial likelihood procedure to get an updated estimate of β. Hence, the parameter β is updated by maximizing[4]

$$L_3(\beta) = \prod_i^n \prod_j^{k_i} \frac{\exp\bigl(\beta' X_i(T_{ij})\bigr)}{\sum_{l=1}^n Y_l(T_{ij})\hat{v}_l \exp\bigl(\beta' X_l(T_{ij})\bigr)}, \tag{17}$$

where $\hat{v}_i = A_i/C_i$. An update of the estimate of the baseline intensity at, say, t_k, is given by

$$\hat{\lambda}_0(t_k) = \frac{\sum_i^n \sum_j^{k_i} dN_{ij}(t_k)}{\sum_{i=1}^n Y_i(t_k)\hat{v}_i \exp\bigl(\beta' X_i(t_k)\bigr)}. \tag{18}$$

In sum, we apply the following alternating EM scheme, which after the initial step iterates between the E-step and the M-step until convergence:

Initial step Use a standard Cox partial likelihood estimation procedure to obtain initial estimates of β and λ_0 from (6) and (7).

M-step Use the current values of θ, β, and λ_0 to compute A_i, C_i and $\hat{v}_i$.

E-step Update the estimates of θ using (15). Update the estimates of β and λ_0 using (17) and (18), respectively.

Klein and Moeschberger (1997) provide on their website a SAS macro for the EM algorithm to estimate the AG model with Gamma distributed unobserved heterogeneity.

Once convergence has been obtained, the covariance matrix of the estimator is computed based on the observed information matrix. Details on the computation of this covariance matrix are given in Appendix B.

4 Illustration

In this section we illustrate our calendar-based interpurchase time model on scanner data for yogurt purchases. We use scanner data from the A.C. Nielsen household scanner panel

[4] Again a Breslow or Efron adjustment should be applied when households purchase at the same time.

data from 1985 to 1988 in Sioux Fall, South Dakota. The data cover a period of 91 weeks. We select those households who are buying only the top brands, that are the brands that are sold frequently enough to build an entire history of the marketing efforts. Furthermore, we restrict the analysis to households who are observed to purchase yogurt at least three times in the observation window, which results in shopping trip information on 598 households. The dataset contains information on price, in-store display, and newspaper feature advertisements at the brand level for each store and each week. The marketing instruments are constant during a week, where the week is defined from Wednesday to but not including Wednesday. We allow households to have multiple purchase occasions during a single week.

For each purchase occasion we know the day and the volume purchased. Furthermore, for each week we know the shelf price (dollars/32oz.) of all brands and which brands are featured or on display. As we do not consider the brand choice we need to aggregate the marketing information over stores and brands. To avoid losing too much information, we use household-specific weights in this aggregation. Following Gupta (1991), we use household-specific volume brand shares to aggregate over brands. Aggregation over stores is carried out using household-specific store weights. Thus, for each household we only use data on the relevant store and brand options. This approach has also been followed by Chib, Seetharaman, and Strijnev (2002) among others. The feature and display variables now represent the percentages of stores (relevant for the specific household) featuring a brand, or having the brand on display. Next to this information on marketing instruments, we use the household size, household income and the volume purchased at the previous purchase occasion. The latter variable is a proxy for inventory.

4.1 Estimation results

We only consider estimation of AG model with random unobserved heterogeneity as discussed in Section 3. This unobserved heterogeneity is time constant and is assumed to have a Gamma distribution with mean 1 and variance θ as in (11), see (Gönül and Srnivasan 1993) for a similar approach in a gap-time approach. As explanatory variables we use household income, household size and the volume purchased at the previous purchase occasion (divided by 32 oz.). The observed volume is decomposed into two variables, that is, time-constant average volume purchased per household and time-varying deviation of this average at each purchase for each household. The first indicates whether a household purchases in large or small amounts (a "regular" trip), while the latter indicates whether the household makes a "fill-in" trip. Because yogurt is only storable for a short period of time, we do not expect that deviations from the average volume will have impact on the consecutive yogurt purchases.

We make a similar decomposition of the actual price (in dollars per 32 oz.) observed. One price variable is the average observed price per household and the other price variable is the regular price minus the average price in each week for each household. The latter is an indicator of a price cut. Two other marketing-mix variables, display and feature, indicate whether brands are on display or are featured in a newspaper. To account for the temporal effect of these marketing instruments, the differences between the current value and the value at the previous purchase of all three variables are also added.

Finally, the duration dependence between consecutive purchases is captured by adding the natural logarithm of the time since the previous purchase. This is equal to assuming a Weibull duration dependence from one purchase to the next. Note that the baseline hazard in our model captures the calendar-time duration dependence, such as seasonal- and day-of-the-week effects, and not the duration dependence between re-occurring purchases. The log time since the previous purchase is one way to incorporate duration dependence in the model. We are, however, not restricted to the Weibull duration dependence. With a slight change of model specification we get an Erlang-2 or Exponential power, which are two popular alternative formulations of the between duration dependence, see Seetharaman and Chintagunta (2003). In sum, the effect that, say, Saturday is a favorite shopping day for many households is reflected in the calendar time duration dependence by peaks at multiples of one week. The possibility that households have shopping trips, say, every week is reflected in a re-occurrence duration dependence with indicators at multiples of one week since the previous purchase.

In Table 1 we present the estimation results with unobserved heterogeneity. The estimated variance of the heterogeneity distribution denoted by θ is 0.41 ($\exp(-0.9)$) and it differs significantly from zero, which indicates that a model without unobserved heterogeneity is not correctly specified. If we consider the volume purchased on the previous purchase occasion, we see that only the household average variable has a significant impact. As expected, any deviation from the household average purchase volume has no significant impact on predicting the repurchase behavior. Household income has a significant negative effect on repurchase intensity, while household size has a positive significant effect. Hence, the lower the household income and the larger the household size the more prone it is to purchase yogurt. The average observed price appears to be an important variable in our model. Households which purchase, on average, more expensive yogurt, are more frequent buyers. A price cut denoted by a deviation from this household-specific average price, has a positive, but insignificant, effect on the repurchase probability. However, an observed negative price difference with the previous purchase has a positive effect on the repurchase probability. Putting yogurt on display in store and or feature it in a newspaper advertisement both have a significant positive direct effect on the repurchase intensity. If we consider the effect of feature and display with respect to the value of feature and display at the previous purchase moment, we see that only the effect of display is significant at a 90% level. The Weibull parameter, which is 0.87 (1-0.1311), indicates a small negative duration dependence. It entails that 10 weeks after the last purchase (with no other changes), the repurchase probability is about 25% lower than just 1 week after the purchase.

4.2 Scenario analyses

In the previous subsection we have discussed the parameter estimates of our model. As already indicated, our model allows forecasting purchases beyond the next purchase, which is not possible with standard gap-time based models. To illustrate the usefulness of our model, we examine in this section the short- and long-run effects on repurchase behavior of three different promotion scenarios. We analyze the effect of a promotion in a single week, which we choose to be week 50. To asses the dynamic impact of the promotions we rely on simulation. We use the estimated model with Gamma distributed heterogeneity to simulate purchases for the next 14 weeks, starting at week 50. All explanatory variables are set at their average value, except for the time-constant household-specific variables (including the implied heterogeneity term). In the first scenario, we introduce a price cut of 33% of the household-specific regular price without feature and/or display support in week 50. In week 51 and beyond we set the marketing-mix variables at their non-promotional value. This implies that after week 50 no display or feature takes place and that the price is equal to the average household-specific price. Note that if a household purchases yogurt in week 50 the difference between the value of the marketing instrument after week 50 and the value at the previous purchase differs from zero until the next purchase. In the second and third scenario we analyze the effect of a feature and display promotion in week 50, respectively.

Our model is very well suited for simulation purposes. In each simulation round we simulate a counting process on a daily interval for each household in our sample. Hence, if we have n households and the horizon is w weeks, each simulation round provides us with a matrix of size $n \times (7w)$ of zeros and ones, where a one corresponds with a purchase of a household on that particular day. From this matrix we can directly derive on a daily basis the simulated number of purchases per household. Furthermore, it is also very convenient to deduce the interpurchase times for each household. The simulation process averages 1000 of these rounds. Then, for each day we calculate the average (simulated) number of purchases across households and the percentage of household making a first or second purchase. All scenarios are compared to a baseline scenario in which the marketing-mix variables are put at their average value in week 50.

Figure 2 shows the effects on the number of purchases in the weeks after the promotion introduction. We see that both display and feature have a strong effect on the number of purchases. The effect of a price cut is relatively small. The effect on the number of purchases in the week after the promotion is negative but it converges to zero soon after. Hence, we here observe the well-known post-promotion dip in interpurchase times, see also van Heerde, Leeflang, and Wittink (2000) among others for a similar pattern in sales. The cumulative effects of display and feature are positive as is shown in Figure 3. A display for a week results in 40 more purchases in the long run than the baseline scenario (for 598 households). A feature increases the number of purchases in the long run by 20 and the effect of a price cut is negligible. Figures 4 shows the percentage of household which made their first purchase after the promotion as a function of time. Figure 5 depicts the same quantities for the second purchase after the promotional week.

5 Conclusion

In this paper we have introduced a statistical model for repeated purchases which considers the whole path of the repurchase history on the calendar timescale. The model is different from the standard gap-time durations models, which are nowadays standard in marketing research. In gap-time models, time is reset to zero after a purchase has been made. Hence, one neglects the purchase history of a household when describing the current interpurchase time. Resetting time to zero implies that the gap-time model is only suited to predict one purchase occasion ahead. In calendar-time based models, which we considered in this paper, time is not reset after a purchase and no information is lost. Furthermore, as we do not reset time, our model can be used to predict more than one purchase ahead. Therefore, the model can also be used to analyze the number and timing of purchases, given a particular scenario of the marketing-mix over time. As the model can be set up in a counting process framework, it is easy to incorporate time-varying covariates and delayed entry.

There are various avenues for further research. The most important of these is to construct a model that simultaneously captures duration and quantity. We also foresee important applications of our approach to describing purchases of durable goods, such as cars and houses.

A Parameter estimation in SAS

Parameter estimation of the counting process models is facilitated by the use of the counting process input style, an option available in SAS and STATA. In this input style each household is represented as a set of rows with time intervals that ends in a purchase or a change in one of the time-varying covariates: (entry time, first change], (first change, second change], ..., (mth change, end of observation window]. The rows of data consists of data observations, each of which contains (fixed) covariate values X, a status indicator $dN_i(t)$ (1 = repurchase, 0 = censored; that is, no repurchase at the end of the time interval), along with the time intervals over which this information applies. If all the covariates are time constant and the household makes only one repurchase before the end of the observation window, that household has only two rows of observation data. The first row contains the information until the repurchase time, (entry time, repurchase time], and the second row contains the information from the repurchase time until the end of the study. In this case the only difference between the two rows of data is that in the first row the status indicator is one, because at the end of the interval the household purchased the good again, while in the second row the status indicator is zero.

If the data also contains time-varying covariates, like a variable that indicates whether yogurt is on display in a particular week, each household has more than two rows of data. Assume, for example, that a household enters the study at 10 weeks, makes its first purchase at $11\frac{2}{7}$ weeks and the second at $15\frac{5}{7}$ weeks. This household falls in income-group 4 and yogurt is on display in week 11 and 14. For each time interval that ends with a purchase or a change in the (time-varying) indicator of display we should have in the counting process input style a separate input row, with the start and end of each interval in the first two columns. Thus the input for this household is coded as in Table 2.

Note that the delayed entry of this household at 10 weeks is accounted for by excluding the lines before week 10. If a household has discontinuous intervals of risk, for example because we know the time the household is away for holiday, the rows of data related to the weeks of holidays are removed from the data. Thus, if the household represented in

Table 2 is on holidays in week 14, this line can be removed from the data.

B The asymptotic covariance matrix

By taking derivatives from the observable log-likelihood (12), we obtain the following components of the observed information matrix:

$$
\begin{aligned}
\frac{-\partial^2 L(\beta,\lambda_0,\theta)}{\partial\lambda_0(t_a)\partial\lambda_0(t_b)} &= \sum_{i=1}^{n} \frac{(1+\theta k_i)Y_i(t_a)\exp\big(\beta' X_i(t_a)\big)Y_i(t_b)\exp\big(\beta' X_i(t_b)\big)}{1+\theta\int_{S_i}^{T_e}\lambda_0(s)\exp\big(\beta' X_i(s)\big)\,ds} \\
&\quad -\sum_{i=1}^{n}\sum_{j=1}^{k_i}\frac{dN_{ij}(t_a)dN_{ij}(t_b)}{\lambda_0^2(t_a)} \\
\frac{-\partial^2 L(\beta,\lambda_0,\theta)}{\partial\lambda_0(t_a)\partial\beta'} &= \sum_{i=1}^{n} \frac{(1+\theta k_i)Y_i(t_a)X_i'(t_a)\exp\big(\beta' X_i(t_a)\big)}{1+\theta\int_{S_i}^{T_e}\lambda_0(s)\exp\big(\beta' X_i(s)\big)\,ds} \\
&\quad -\sum_{i=1}^{n}\frac{\theta(1+\theta k_i)Y_i(t_a)\exp\big(\beta' X_i(t_a)\big)\int_{S_i}^{T_e}\lambda_0(s)X_i'(s)\exp\big(\beta' X_i(s)\big)\,ds}{\Big[1+\theta\int_{S_i}^{T_e}\lambda_0(s)\exp\big(\beta' X_i(s)\big)\,ds\Big]^2} \\
\frac{-\partial^2 L(\beta,\lambda_0,\theta)}{\partial\lambda_0(t_a)\partial\theta} &= \sum_{i=1}^{n} \frac{k_i Y_i(t_a)\exp\big(\beta' X_i(t_a)\big)}{1+\theta\int_{S_i}^{T_e}\lambda_0(s)\exp\big(\beta' X_i(s)\big)\,ds} \\
&\quad -\sum_{i=1}^{n}\frac{(1+\theta k_i)Y_i(t_a)\exp\big(\beta' X_i(t_a)\big)\int_{S_i}^{T_e}\lambda_0(s)\exp\big(\beta' X_i(s)\big)\,ds}{\Big[1+\theta\int_{S_i}^{T_e}\lambda_0(s)\exp\big(\beta' X_i(s)\big)\,ds\Big]^2}
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{-\partial^2 L(\beta,\lambda_0,\theta)}{\partial\beta\partial\beta'} &= \sum_{i=1}^{n} \frac{(1+\theta k_i)\int_{S_i}^{T_e}\lambda_0(s)X_i(s)X_i'(s)\exp\big(\beta' X_i(s)\big)\,ds}{1+\theta\int_{S_i}^{T_e}\lambda_0(s)\exp\big(\beta' X_i(s)\big)\,ds} \\
&\quad -\sum_{i=1}^{n}\frac{\theta(1+\theta k_i)\int_{S_i}^{T_e}\lambda_0(s)X_i(s)\exp\big(\beta' X_i(s)\big)\,ds\int_{S_i}^{T_e}\lambda_0(s)X_i'(s)\exp\big(\beta' X_i(s)\big)\,ds}{\Big[1+\theta\int_{S_i}^{T_e}\lambda_0(s)\exp\big(\beta' X_i(s)\big)\,ds\Big]^2} \\
\frac{-\partial^2 L(\beta,\lambda_0,\theta)}{\partial\beta\partial\theta} &= \sum_{i=1}^{n} \frac{k_i\int_{S_i}^{T_e}\lambda_0(s)X_i(s)\exp\big(\beta' X_i(s)\big)\,ds}{1+\theta\int_{S_i}^{T_e}\lambda_0(s)\exp\big(\beta' X_i(s)\big)\,ds} \\
&\quad -\sum_{i=1}^{n}\frac{(1+\theta k_i)\int_{S_i}^{T_e}\lambda_0(s)\exp\big(\beta' X_i(s)\big)\,ds\int_{S_i}^{T_e}\lambda_0(s)X_i(s)\exp\big(\beta' X_i(s)\big)\,ds}{\Big[1+\theta\int_{S_i}^{T_e}\lambda_0(s)\exp\big(\beta' X_i(s)\big)\,ds\Big]^2}
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{-\partial^2 L(\beta,\lambda_0,\theta)}{\partial\theta^2} \quad &= \quad \sum_{i=1}^{n}\frac{k_i}{\theta^2}+\frac{\psi'(\frac{1}{\theta})-\psi'(\frac{1}{\theta}+k_i)}{\theta^4}+\frac{2\Big(\psi(\frac{1}{\theta})-\psi(\frac{1}{\theta}+k_i)\Big)}{\theta^3} \\
&\quad + \sum_{i=1}^{n}\frac{2}{\theta^3}\ln\Big[1+\theta\int_{S_i}^{T_e}\lambda_0(s)\exp\big(\beta' X_i(s)\big)\,ds\Big] \\
&\quad - \sum_{i=1}^{n}\frac{2\int_{S_i}^{T_e}\lambda_0(s)\exp\big(\beta' X_i(s)\big)\,ds\int_{S_i}^{T_e}\lambda_0(s)X_i(s)\exp\big(\beta' X_i(s)\big)\,ds}{\theta^2+\theta^3\int_{S_i}^{T_e}\lambda_0(s)\exp\big(\beta' X_i(s)\big)\,ds} \\
&\quad - \sum_{i=1}^{n}(1/\theta+k_i)\left[\frac{\int_{S_i}^{T_e}\lambda_0(s)\exp\big(\beta' X_i(s)\big)\,ds\int_{S_i}^{T_e}\lambda_0(s)X_i(s)\exp\big(\beta' X_i(s)\big)\,ds}{1+\theta\int_{S_i}^{T_e}\lambda_0(s)\exp\big(\beta' X_i(s)\big)\,ds}\right]^2,
\end{aligned}
$$

where $\psi'(\cdot)$ is the second derivative of the log Gamma function, that is, $\psi'(x)=\partial^2\Gamma(x)/\partial x^2$.

Table 1: Parameter estimates for the AG model with Gamma distributed unobserved heterogeneity. Standard errors in parentheses

variable	estimate	standard error
household income	-0.030	(0.001)
household size	0.064	(0.002)
volume previous purchase (compared to household average)	-0.066	(0.038)
volume previous purchase (household average)	0.250	(0.010)
price (compared to household average)	-0.167	(0.142)
price (household average)	0.626	(0.007)
display	0.885	(0.115)
feature	0.539	(0.093)
price difference (compared to previous purchase)	-0.173	(0.079)
display difference (compared to previous purchase)	0.256	(0.140)
feature difference (compared to previous purchase)	0.172	(0.130)
log of time since previous purchase	-0.131	(0.004)
$\ln(\theta)$	-0.889	(0.067)
value of the log-likelihood function	4170.09	

Table 2: Example of counting process input style

Start	end	status	income	display
10	11	0	4	0
11	$11\frac{2}{7}$	1	4	1
$11\frac{2}{7}$	12	0	4	1
12	14	0	4	0
14	15	0	4	1
15	$15\frac{5}{7}$	1	4	0
$15\frac{5}{7}$	16	0	4	0
.	.	.	.	.

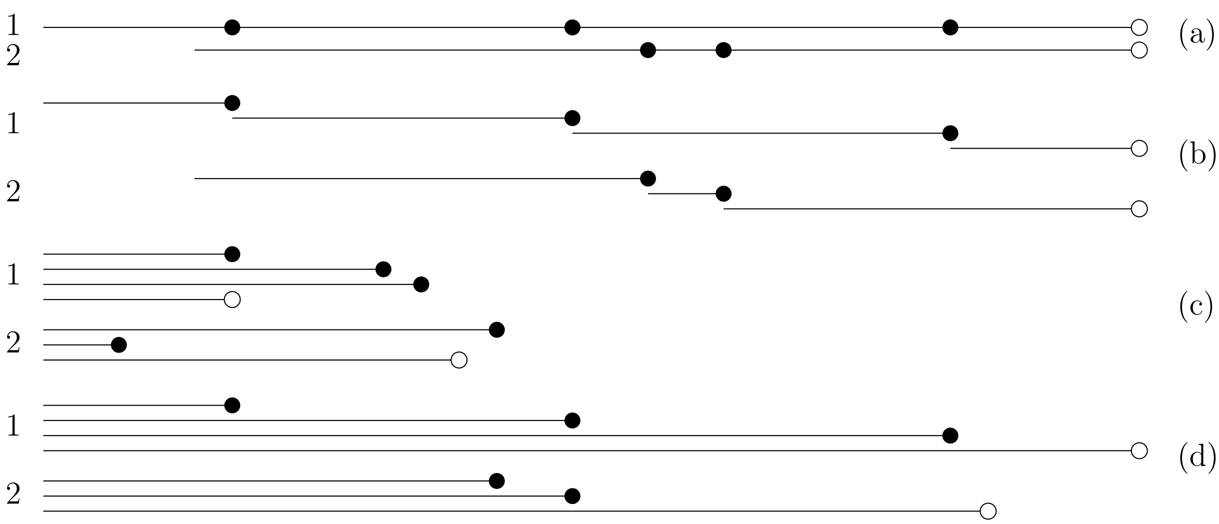

Figure 1: Illustration of different timescales: (a) Purchase history of two households; (b) in calendar time; (c) in gap time; (d) in total time. (a closed circle marks a purchase and an open circle a censored observation)

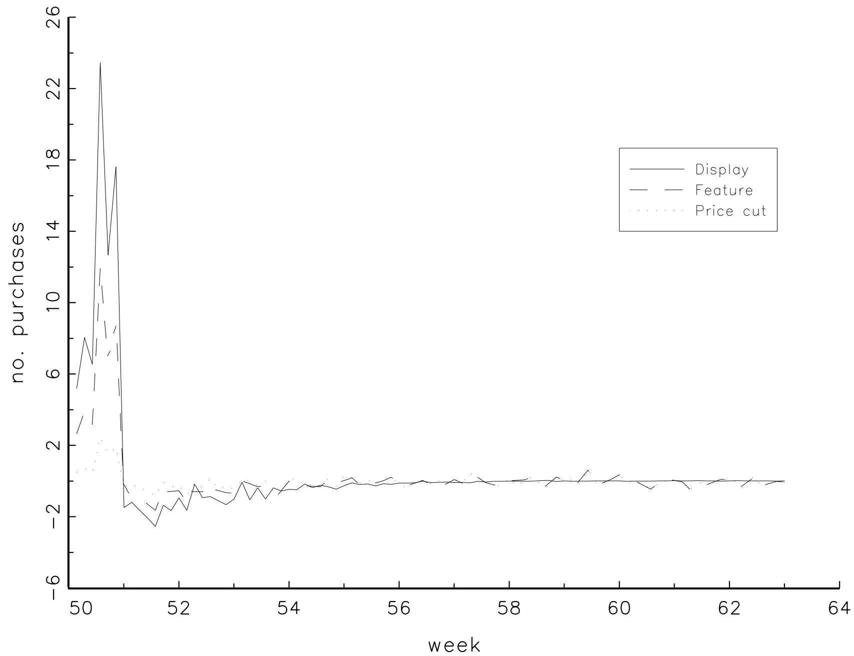

Figure 2: Effect of promotion instruments in week 50 on number of purchases

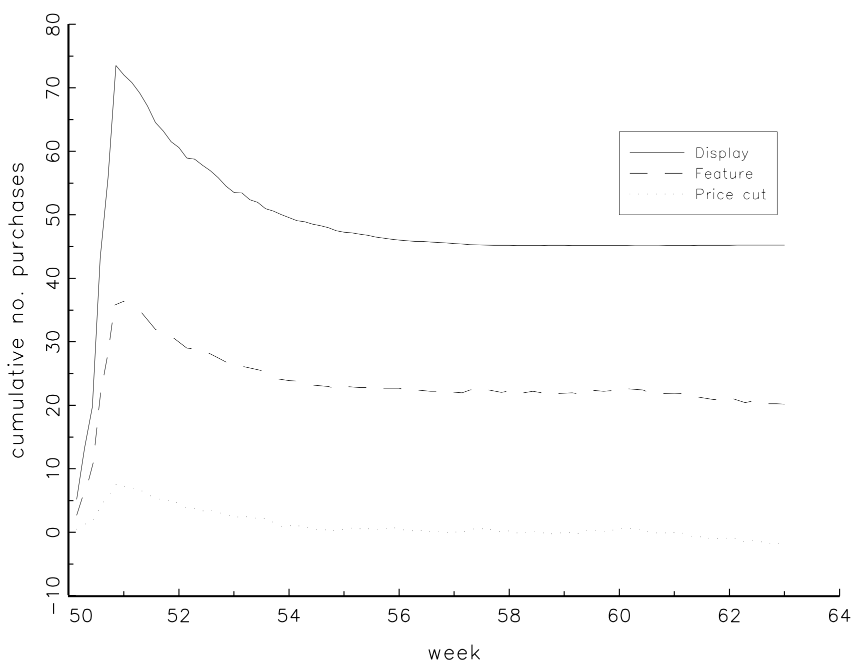

Figure 3: Cumulative Effect of promotion instruments in week 50 on number of purchases

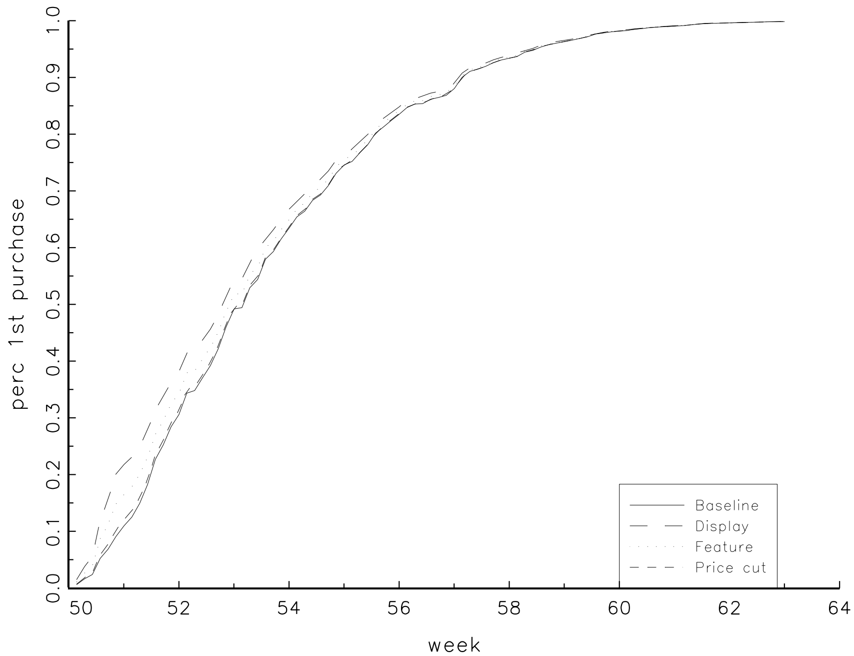

Figure 4: Effect of promotion instruments in week 50 on timing of first repurchase

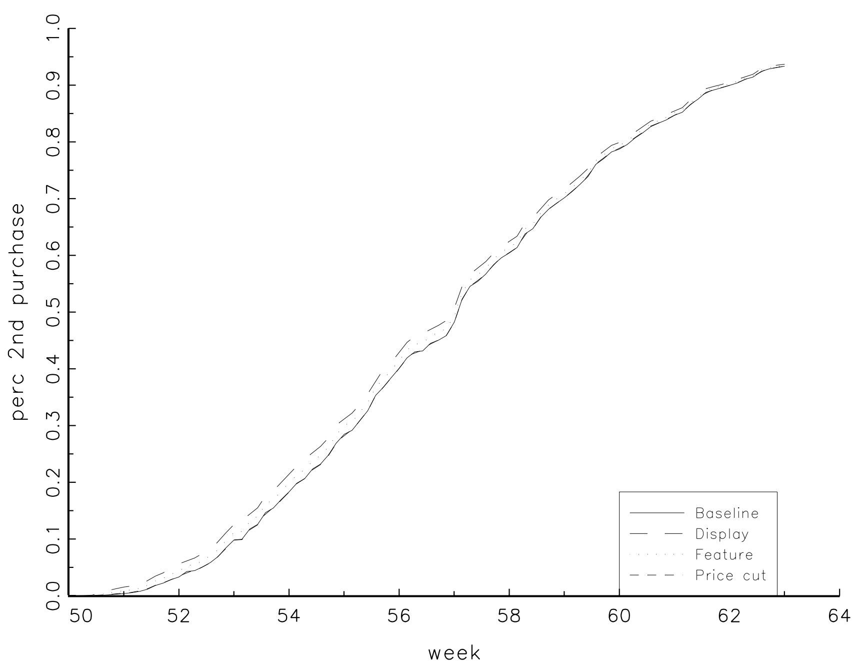

Figure 5: Effect of promotion instruments in week 50 on timing of second repurchase

References

Andersen, P. K., O. Borgan, R. D. Gill, and N. Keiding (1993). *Statistical Models Based on Counting Processes*. New York: Springer–Verlag.

Andersen, P. K. and R. D. Gill (1982). Cox's regression model for counting processes: A large sample study. *Annals of Statistics 10*, 1100–1120.

Breslow, N. (1972). Discussion on 'regression models and life-tables' (by D.R. Cox). *Journal of the Royal Statistical Society: Series B 34*, 187–220.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics 30*, 89–99.

Chib, S., P. B. Seetharaman, and A. Strijnev (2002). Analysis of multicategory purchase incidence decisions using IRI market basket data. In P. H. Franses and A. L. Montgomery (Eds.), *Advances in Econometrics*, Volume 16, pp. 57–92. Elsevier Science.

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

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

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B 39*, 1–38.

Duchateau, L., P. Janssen, I. Kezic, and C. Fortpied (2003). Evolution of recurrent ashtma event rate over time in frailty models. *Applied Statistics 52*, 355–363.

Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association 72*, 557–565.

Gönül, F. and K. Srnivasan (1993). Consumer purchase behavior in a frequently bought prodcut category: Estimation issues and managerial insights from a hazard function model with heterogeneity. *Journal of the American Statistical Association 88*, 1219–1227.

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

Helsen, K. and D. C. Schmittlein (1993). Analyzing duration times in marketing: Evidence for the effectiveness of hazard rate models. *Marketing Science 11*, 395–414.

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

Kelly, P. J. and L. Y. Lim (2000). Survival analysis for recurrent event data: An application to childhood infectious diseases. *Statistics in Medicine 19*, 13–33.

Klein, J. P. (1992). Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics 48*, 795–806.

Klein, J. P. and M. L. Moeschberger (1997). *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer–Verlag.

Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrica 47*, 939–956.

Lawless, J. F. (1987). Regression methods for poisson process data. *Journal of the American Statistical Association 82*, 808–815.

Lin, D. Y. and L. J. Wei (1989). The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association 84*, 1074–1078.

Nielsen, G. G., R. D. Gill, P. K. Andersen, and T. A. I. Sørensen (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics 19*, 25–43.

Oakes, D. A. (1992). Frailty models for multiple event times. In J. P. Klein and P. K. Goel (Eds.), *Survival Analysis: State of the Art*, pp. 371–379. Dordrecht: Kluwer.

Prentice, R. L., B. J. Williams, and A. V. Peterson (1981). On the regression analysis of multivariate failure time data. *Biometrika 68*, 373–379.

Rondeau, V., D. Commenges, and P. Jolly (2003). Maximum penalized likelihood estimation in a gamma–frailty model. *Lifetime Data Analysis 9*, 139–153.

Seetharaman, P. B. and P. K. Chintagunta (2003). The proportional hazard model for purchase timing: A comparison of alternative specifications. *Journal of Business & Economic Statistics 21*, 368–382.

Therneau, T. and P. Grambsch (2000). *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag.

van Heerde, H. J., P. S. H. Leeflang, and D. R. Wittink (2000). The estimation of pre- and postpromotion dips with store-level scanner data. *Journal of Marketing Research 37*, 383–395.

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

Wei, L. J., D. Y. Lin, and L. Weissfeld (1989). Regression analysis of multivariate failure time data by modeling marginal distributions. *Journal of the American Statistical Association 84*, 1065–1073.

Winkelmann, R. (1995). Duration dependence and dispersion in count data models. *Journal of Business & Economic Statistics 13*, 467–474.