When Are Your Customers Active and is their Buying Regular or Random?

An Erlang Mixture State-Switching Model for Customer Scoring

Joachim Büschken
School of Management
Catholic University of Eichstätt-Ingolstadt, Germany
jb@wfi.edu

and

Shaohui Ma
School of Economics and Management
Jiangsu University of Science and Technology, China
msh@tju.edu.cn

March, 2012

Abstract

Scoring customers with regard to the expected number of their future transactions is of fundamental interest to direct marketers. For the purpose of customer scoring, a variety of models have been developed, based on the negative binomial distribution. Extensions of such models allow for customer defection (“buy until you die”). We extend customer scoring models in two ways: (1) We assume that customers switch between an active state in which they buy and an inactive state in which they do not purchase. Switching between states can occur in both directions, implying that defection is temporary. (2) We allow for heterogeneity among households with regard to the regularity of purchasing. We achieve this by modeling purchasing in the active state as a mixture of Erlang-distributed interpurchase times. Our model nests the BG-NBD model and other variants of the NBD as special cases. We develop MCMC simulation methods for our model and apply it to five different data sets, comparing the results to several benchmark models. Our results imply that it is often erroneous to assume that customers “buy until they die”. Instead, the results support the assumption of transient switching between an active and inactive state. We show that, when the model does not account for transient switching, customer scoring is biased. We discuss both the theoretical and managerial implications of our results, and potential areas for future research.

1 Shaohui Ma acknowledges the ongoing support of the National Natural Science Foundation of China (NSFC) under grants no. 70871057, 71171100. Panel data were provided by GfK e.V. Nürnberg (Germany) and the CDNow data by Bruce Hardie (LBS). This support is gratefully acknowledged.
When Are Your Customers Active and is their Buying Regular or Random?
An Erlang Mixture State-Switching Model for Customer Scoring

1. Introduction

The successful execution of direct marketing campaigns requires firms to predict the future transactions of individual customers. Questions such as (1) How many transactions can be expected from Customer X in the next year? or (2) What is the probability of Customer X being an active customer? are of fundamental interest to direct marketers. Increasingly, such questions are considered in domains other than the mail-order business. This may occur in financial services firms where direct marketing activities such as targeted mailings, or direct interactions with customers through call-center agents, are the norm. For example many brick-and mortar retailers have developed the ability to collect transaction data at the individual household level, through the use of loyalty programs. Other examples include internet-based firms which require customers to identify themselves for every purchase. Such firms can easily collect household-level purchase data and engage in direct marketing based on this information.

An interesting issue arises when firms consider questions such as: Should one include the option of customers defecting in modeling purchasing behavior? And if so, should one think of defection as something “terminal”, with a defected customer never returning (or at least, not returning within a reasonable amount of time)? Or should one think of defection as temporary? Regarding defection as terminal has strong implications. When is it reasonable to assume that to a defected customer will never return? It may be a reasonable assumption for a bank that lost a homeowner to another mortgage lender. However, for a retailer or consumer brand manager, whose customers have relatively low switching costs, it may not be reasonable to make this assumption.
Many customer scoring models assume that defection is indeed “terminal”. This applies to models based on the Pareto NBD framework (Schmittlein et al. 1987), which are considered as benchmark models for customer scoring (Wübben and Wangenheim 2008). The purpose of this paper is to relax the assumption of terminal defection and to allow for customers who defect temporarily. From the perspective of this paper, defection occurs when customers become temporarily inactive. In this state, their purchase rate is zero. They may return subsequently to an active state in which their purchase rate is positive. Our state-switching model allows making predictions about when inactive customers will become active again. It treats “buy-until-you-die” models as a special case, since we allow the recovery rate of inactive customers to become zero. It nests various NBD-type models which assume that customers are always active, by allowing for the probability of becoming inactive to be zero. Thus, the model we propose presents a unifying framework for various types of customer scoring approaches. The paper is organized as follows. Section 2 outlines prior research relevant to our model. Section 3 develops the proposed mixture of Erlang state-switching model. We present results from a simulation study which demonstrate that the proposed model is identified. In Section 4, we apply our model to five different transaction datasets. Section 5 discusses the empirical results. Section 6 discusses the key findings and concludes with possible areas for future research.

2. Prior research

Prior research relevant to our modeling approach falls into three categories: the modeling of consumer interpurchase times, research on state-switching models and research on the regularity of consumer purchasing behavior. We address these areas in turn.
The workhorse model for customer scoring is the Pareto NBD (Schmittlein et al. 1988), which assumes that the number of transactions across customers follows the NBD distribution. The waiting time until customers defect follows a Pareto distribution. An implication of the Pareto NBD is that, after defection, customer have a zero probability of buying again. Wübben and Wangenheim (2008) compare predictions generated by the Pareto NBD to those from naïve heuristics and find that the Pareto NBD does not always generate more accurate predictions than simpler approaches.

Researchers have advanced the “NBD plus defection” model in various ways. Schmittlein and Peterson include the monetary value of transactions through a normal distribution (1994). Fader et al. (2005) introduced the BG NBD model, a simpler version of the Pareto NBD, which ties (terminal) defection to purchase occasions. Jareth et al. (2011) extend the BG NBD by allowing for defection to be periodic and, thus, independent of the purchase process. When periodicity approaches infinity or zero, the Pareto NBD and the “no death/always active” NBD model arise as limiting cases, respectively. Batislam et al. (2007) modify the BG NBD model to allow for zero repeat purchases. Abe (2009) applies a hierarchical Bayes approach to the Pareto NBD, that enables the estimation of individual-level quantities of interest.

These extensions of the Pareto NBD are very useful for customer scoring, as long as we assume that, after defection, customers will not return. We call this “terminal” defection. But what if customers become temporarily inactive after which they become active again? The “buy-until-you-die” assumption precludes this possibility. In this situation, instead of assuming terminal defection, we wish to model defection as a temporal, non-absorbing state. This would be more consistent with the type of buying behavior typically observed in always-a-share markets.
Prior research on retention in always-a-share markets has addressed purchasing dynamics in various ways. Important contributions have resulted from the analysis of purchasing in always-a-share markets, where switching is most likely to be transient. Early examples of modeling transient switching between recency/frequency states are presented by Bitran and Mondschein (1996), Gonul and Shi (2000) and Pfeifer and Carraway (2000). Rust et al. (2004) model transition probabilities between brands across time. Recency states and brand choice are observable, giving rise to switching matrices directly derived from observed data. Allenby et al. (1999) model switching between unobserved purchase rates. Customers display unobserved time-variant purchase rates, associated with different levels of activity. In the “super-active” state, customers purchase in a clockwise fashion and in every period. In the “active” state, customers buy less often and less regularly. In the “inactive” state, customers’ interpurchase times become longest and their buying becomes purely random. The observed interpurchase times at the individual level are assumed to be generated by a mixture of generalized gamma distributions with customer and state-specific purchase rates. This modeling approach introduces dynamics to individual-level purchasing behavior, as customers can move from one activity level to another with every purchase.

A different form of purchasing dynamics through state-switching is introduced in the model proposed by Ma and Büschken (2011). In this model, customers switch between an inactive and an active state, giving rise to an interrupted purchasing process. In the inactive state, customer purchase rates are fixed at zero. Thus, when inactive, no purchase can be made. In the active state, interpurchase times are distributed exponentially, given individual-level purchase rates. The observed interpurchase times may contain waiting times for the customer to switch from the inactive back to the active state (“recovery”). When the recovery rate approaches zero,
the waiting time until a customer switches back to the active state approaches infinity and the model approaches the BG NBD model.

**Regularity of Buying**

Observed interpurchase times may not be exponentially distributed (Chatfield and Goodhardt 1973, Wheat and Morrison 1990, 1994). Instead, purchases may be more regular than suggested by the exponential distribution. Distributions of interpurchase intervals with a non-zero mode are an indicator of greater regularity. Non-random buying has motivated researchers to model interpurchase times in ways that allow for greater buying regularity. One way to achieve this is to use the Erlang-distribution (e.g. Morrison and Schmittlein 1988). The Erlang distribution is a special case of the gamma distribution with a positive, integer-valued shape parameter ($k$). The exponential distribution is a special case of the Erlang with $k=1$. When the shape parameter of the Erlang approaches infinity, purchasing approaches a “clockwise” pattern. Researchers have developed methods to analyze purchase regularity at the individual level (e.g. Wheat 1987). Empirical results using such methods indicate that the Erlang distribution with $k=2$ is a good approximation of consumer buying behavior. Morrison and Schmittlein (1988) develop the Condensed NBD model on that basis.

More recently, researchers have successfully used the generalized gamma distribution to model interpurchase times (Allenby et al. 1999). The generalized gamma distribution nests the exponential, the Erlang, the gamma and the Weibull distribution. The shape parameter of the gamma distribution is not restricted to integer values and estimated from the data. This allows for purchasing to be more regular than random. Using transaction data of customers of an investment brokerage firm, Allenby et al. (1999) find that in the super-active state, $k \approx 50$. In the
active state, $k=7$. This indicates that, in states with higher activity levels, individual transactions are more regular than random.

Towards a better understanding of purchase timing

Our model builds on the research outlined above and its contribution is twofold. Firstly, we consider the possibility that customers defect temporarily after a purchase and then resume an active state. In the inactive state, the purchase rate is zero. This approach introduces dynamics to the purchase rate (0 when inactive, positive when active). A purchase rate fixed at 0 in the inactive state captures inactivity instead of a lower level of activity. In the Allenby et al. (1999) model, purchase rates cannot become zero, as customers are always assumed to be active. In our model, we assume that customer may become inactive, and that the time needed to switch back to an active state is governed by the rate of a separate process, the unobserved “recovery process”. This assumption departs from that of “always-active” customers in the Allenby et al. model. The switching in both directions (active to inactive, inactive to active) also differentiates our model from those with the “buy-until-you-die” assumption.

Secondly, we account for the possibility that purchasing at the individual level can be more regular than random. We achieve this through a mixture distribution where interpurchase times in the active state are drawn from a mixture of Erlang($k$) distributions. Although we choose a specific number of components and of values of $k$ in the empirical application of our model, our model is developed flexibly and is therefore not limited to a specific number of components with specific values of $k$. The mixture of Erlang processes is a novel perspective to modeling purchase timing. We assume that a sample of customers represents a mix of random and regular buyers in a priori unknown proportions. Because we do not wish to “fix” customers a priori to being
“always active” or “buying until they die”, we model this mixture in the context of switching between an active and inactive state, by restricting the mixture to active phases.

What makes our approach innovative? If switching is transient and consumers have periods of inactivity instead of defecting, it is of interest to obtain information about when inactive customers resume the active state. Our model generates information about the rate at which inactive customers become active. For direct marketers, it is useful to know the recovery rate and to predict when the active state is resumed. Since our model contains terminal defection as a limiting case, the question of whether households defect temporarily or permanently is treated as an empirical rather than a conceptual issue.

Ultimately, our model is intended to provide a better understanding of purchase timing. We assume that observed interpurchase times contain unobserved “downtimes”. During downtimes, consumers have a purchase rate of zero. A priori, it is reasonable to assume the existence of downtimes in markets such as consumer durables, where purchases may be followed by a lengthy period of usage and purchase inactivity. Other examples are non-perishable food products, on which customers stock up and become inactive until stocks are depleted and a new shopping trip is necessary. A model that does not differentiate between downtimes and active times and assumes that customers are always active until they defect, underestimates the purchase rate of customers, regardless of their level of activity. An underestimation of purchase rates would bias customer scoring.

3. Model Development

Assumptions

The proposed model, which we call the Erlang mixture state-switching model (EMS), is based on the following assumptions:
1. At any time $t$, a customer can be active or inactive. While inactive, customers make no transactions. While active, waiting times between transactions follow a mixture of $Q$ Erlang distributions of order $k_q$, individual-level purchase rates $\lambda_i$ and component weights $\varphi_q$.

$$t_{ij|active} \sim \sum_q \varphi_q \text{Erlang}(k_q, \lambda_i)$$

2. After any purchase, customer $i$ becomes inactive with probability $p_i$. While inactive, the waiting time until customer $i$ becomes active again (we call this the “recovery waiting time”) is exponentially distributed, given the individual-level recovery rate $\gamma_i$.

3. Heterogeneity in transaction rates across customers follows a gamma distribution with shape parameter $a_\lambda$ and scale parameter $b_\lambda$:

$$\lambda_i \sim \text{Gamma}(a_\lambda, b_\lambda)$$

4. Heterogeneity in recovery rates across customers follows a gamma distribution with shape parameter $a_\gamma$ and scale parameter $b_\gamma$:

$$\gamma_i \sim \text{Gamma}(a_\gamma, b_\gamma)$$

5. Heterogeneity in dropout probabilities across customers follows a beta distribution with shape parameter $a_\beta$ and scale parameter $b_\beta$:

$$p_i \sim \text{Beta}(a_\beta, b_\beta)$$

6. Transaction rates, recovery rates and dropout probabilities vary independently across customers.

In Appendix A, we outline a Bayesian estimation approach for our model. Because of its flexible set-up (e.g., the individual recovery rate can approach 0, implying terminal defection), it serves as a unifying framework for alternative customer scoring models, based on purchase events (see Table 1):
For $q=1$, $k=1$ and $\gamma = 0$ our model represents a hierarchical Bayes version of the BG NBD model (Fader et al. 2005). This is because, when $\gamma = 0$, the waiting time until recovery approaches infinity and defection becomes terminal.

- For $q=1$ and $k=1$, we obtain the state-switching model proposed by Ma and Büschken (2011).

- For $q=1$, $p = 0$ and $k=1$, we obtain the NBD model in which customers are always active, for $k=2$, we obtain the condensed NBD (Morrison and Schmittlein 1994).

[Table 1]

The above demonstrates that the EMS model nests several models in the literature. The hierarchical set-up enables us to capture heterogeneity among households in two ways. Firstly, it captures heterogeneity among households with regard to purchase regularity. The waiting times in the active state are assumed to be draws from a mixture of Erlang distributions with household-level purchase rates. Secondly, we capture heterogeneity in the defection/recovery process by modeling defection probabilities and recovery rates as random effects, estimating these quantities using individual-level purchase data, while pooling information across subjects.

**Simulation Study**

We demonstrate model identification and the functionality of our MCMC sampler for the EMS model through a simulation study. The study is based on 250 households for each of two

---

2 Readers might object to this statement. A hierarchical implementation of the (BG) BND explicitly separates the individual-level likelihood from the distributions of heterogeneity and estimates the parameters of both, using Bayes’ Theorem. In contrast, in the BG NBD, the likelihood function is obtained by integrating over the individual-level parameters. When we refer to other models (such as the BG NBD) as being limiting cases of our model, we depart from this fact and refer only to the underlying assumptions of the model.
Erlang($k$) mixture components with $k \in \{1,2\}$. We draw the true household-level quantities from the following hyperprior distributions:

\[ \lambda_i \sim \text{Gamma}(10,10) \]
\[ \gamma_i \sim \text{Gamma}(10,100) \]
\[ p_i \sim \text{Beta}(12,100) \]
\[ k_i \sim \text{Bernoulli}(0.5,0.5) \]

Given $\theta_i$, we generate interpurchase times from simulated recovery and transactional waiting times for $T=250$ periods. Given this simulation set-up, the median of the number of transactions and $T_0$ (total recovery waiting times) across individuals are 92 and 103, respectively. This set-up implies that the average individual is inactive for about 40% of the observation period. For estimation purposes, we use weakly informative priors (see appendix A for the prior set-up) and start the MCMC sampler at arbitrary values. Starting the MCMC sampler at different values leads to estimation results that are almost identical to the results reported in Table 2.

[Table 2]

The results of the simulation study summarized in Table 2 verify that our model is identified and that our MCMC sampler works. We recover the hyperparameters up to the level of certainty indicated by the estimated posterior distribution. Table 2 also reports the MAPE for the individual-level quantities. These parameters can be recovered at reasonable levels of accuracy, given the simulation set-up.

4. Empirical Application

Data

We investigate the empirical performance of our model using five different datasets. Table 3 describes the data. We use left-censored panel data (from a food discounter retail chain
and a drugstore chain) as well as uncensored data, containing an acquisition date (consumer electronics, purchase/renting of video equipment, CD) and household-specific observation periods. The food discounter data and the drugstore chain set were obtained from a single consumer panel data set. This panel data set contains the purchases of 30,000 German households over a period of 2 years (June 2007 to June 2009). From this panel data set, we first excluded all households from the analysis, which did not continuously report their purchases throughout the observation period (about 20% did not report continuously and were therefore excluded from the analysis). We then drew two random samples of 2,000 households each, one sample for each retail company. Thirdly, we selected those households from the two random samples which made at least one transaction with the focal retailer during the observation period. This step identifies households which are actually customers of the focal retailers. The selection procedure results in different sample sizes for the two panel data sets (see Table 3). For the food discounter and the drugstore data, we use data for the first 80 weeks to calibrate our model and the last 25 weeks to validate predictions.

Additionally, we obtained transaction data from a European consumer electronics retailer. This firm is the largest consumer electronics retailer in its home country. In this data set, household-specific observation periods are not left-censored. The longest observation period is 6 years and the shortest, 4 months. The fourth data set used in this study is that from Ma and Büschken (2011). This data set consists of 401 customers of a professional video equipment sales and rental company in China. The fifth data set are the CDNow data (e.g. Fader et al. 2005, 2006; Abe 2009). For each data set, we use the number of customers indicated in Table 3 for calibration and out-of-sample predictions.

[Table 3]
Table 3 presents descriptive statistics for the five data sets. From Table 3, it is clear that transaction patterns differ greatly among the four firms. The median number of transaction is 30 for the food discounter whereas and 7 for the drugstore chain. The standard deviation of the number of transactions at the food discounter is 40.2, while for the drugstore, it is considerably smaller at 15.9. Thus, heterogeneity among households in transaction activity at the food discounter is much larger than purchases made from the drugstore chain. The distribution of the number of transactions across households is markedly skewed to the left for both firms, as indicated by the difference between median and mean. An analysis of the interpurchase times across households reveals that the median interpurchase time for the food discounter is 1 week (mean 1.9 weeks), which implies that a significant number of households buys at short intervals. The standard deviation of the interpurchase times is 4.3. Average interpurchase times at the drugstore chain are roughly three times longer (5.9 instead of 1.9 for the food discounter). The standard deviation of interpurchase times for the drugstore is 10.4. However, the coefficient of variation of interpurchase times for the drugstore chain (1.8) is smaller than for the food discounter (2.3).

We observe fewer transactions and longer interpurchase times for all other data sets. For the consumer electronics retailer and the video equipment company, the median number of transactions is 1, indicating that most customers make only one repeat purchase after their initial transaction. Consumer electronics are purchased at long intervals (mean interpurchase time of 33 weeks). This is consistent with the sporadic buying of expensive consumer durables such as TVs or kitchen appliances. The video equipment company caters mainly for professional video producers, many of whom conduct transactions in a more regular fashion, as indicated by a mean interpurchase time of 7 weeks.
The CDNow data have been used in multiple papers that model transaction events (e.g. Fader et al. 2005, Abe 2009). The characteristics of this data set are markedly different from the other four data sets. The majority of customers in the CDNow data set makes no repeat purchase (1,411 out of 2121 customers). The mean observation period is the shortest (33 weeks). This data set presents a situation in which relatively little information is available at the customer level. The method of information pooling across customers then becomes critical.

We measure recency in the usual way, as the time between the end of the observation period and the last transaction. Median recency for the food discounter is 1.7 weeks and 6.3 weeks for the drugstore chain. Average recency is by an order of magnitude larger: 7.7 and 16.4, respectively, for food and the drugstore data. This implies that the distribution of recency across households in both data sets is heavily skewed. Average recency at the consumer electronics retailer is 52 weeks, the highest among all data sets. The standard deviation of recency is 55.7. Apparently, some customers defect terminally when recency approaches the length of the observation period. The average recency in the CDNow data set is 26 weeks. This value is driven by customers with no repeat purchase, where recency equals the length of the observation period. Average recency is not readily comparable across the five data sets, since the length of the observation period varies (Table 3). It is therefore useful to compare average recency normalized by the average length of the observation period, given the data set. This score indicates average recency by average unit of observation time. Average recency in the food discounter data is 10% of the average (in this case constant) observation period. In the video tech data, this score is 40%. In the CDNow data, it is highest (80%), implying that, on average, customer were inactive 80% of the observed time. It seems reasonable to conclude that in the CDNow data, given that the average customer has remained inactive for most of the observation time, many of these customers have defected terminally. In the food data set, that is clearly not the case. In
conclusion, our data sets span a wide range of purchase timing patterns and, presumably, a correspondingly wide range of potential unobserved activity/inactivity patterns.

5. Results

Model comparison

The last column of Table 1 describes the mixture models that we apply to the five datasets. We use $k \in \{1,2\}$ to allow for a mixture of households with random and more regular buying. We do not fit all the models described in Table 1, as this would add significant clutter to the empirical results. The mixture nests single component models with $k=1$ or $k=2$. We estimate the models using the approach outlined in Appendix A. For the restricted models, we fix parameters at the values indicated in Table 1. We use the log-marginal density, given the model, as a measure of comparison. By definition, the log-marginal density penalizes over-parameterized models. In-sample fit and predictive fit results are summarized in Table 4. The predictive fit is the likelihood of observing the hold-out data, given the group-level estimates obtained from the calibration sample.

[Table 4]

Across the models and data sets, we find that the EMS outperforms the other models in terms of in-sample fit and predictive fit, with one exception only. This exception is the CDNow data, where the predictive fit of the Buy-Until-You-Die (BTYD) model is significantly better than that of the state-switching model (log marginal density of -1,302 for BTYD compared to -1,467 for the EMS). This result has interesting implications. The CDNow data have been used in various applications of stochastic customer scoring models. It seems that these applications have favored the BTYD assumption.
The goal of customer scoring is to predict the number of transactions in the future. We compute the conditional expectation of the number of future transactions, $E(Y(t)|\theta, k=1)$, using the results from Ma and Büschken (2011). For $k=2$, a closed-form expression for $E(Y(t))$ is not available. In Appendix B, we present a way to compute this quantity at the desired level of accuracy. This is a new result. At each iteration of the MCMC chain, using the expression for $E(Y_i(t)|\theta, k=2)$, we obtain a prediction of the number of transactions in the validation period. We average these predictions across the MCMC iterations for each customer and then compute the mean squared error (MSE) at the customer level. This quantity is then averaged across customers. Table 5 summarizes the results from comparing customer-level predictions to data from the validation period. We do not report results for the video equipment company because the average observation period is too short to set data aside for a meaningful validation of customer-level predictions.

[Table 5]

Table 5 shows that the state-switching model outperforms the others in terms of prediction error for the food discounter, the drugstore and the CDNow data. Predictions obtained from the BTYD model outperform the state-switching model for the consumer electronics data set (MSE of 0.99 compared 1.04 for state-switching). One reason for this result could be that the length of the prediction period is too short for the slow recovery of customers in this data set to have an effect on predictions. If the prediction period is short and $\gamma$ is small, predictions from state-switching become very similar to those obtained from the BTYD model.

In Table 5, we also report the MSE obtained from a naïve prediction which averages transactions per week for each household during the observation period, and then extrapolates this buying pattern for the validation period. Interestingly, the naïve model outperforms the BTYD and the always active model for the food discounter data. The naïve approach always
performs worse than the state-switching model, for some data sets by an order of magnitude suggesting that naïve heuristics don’t “get it right” (Wübben and Wangenheim 2008). Overall, we observe a tendency for model-based individual-level predictions to outperform the naïve model-free approach and for the state-switching model to outperform the simpler models.

**Estimation Results: Firm Level**

Table 6 summarizes the estimation results obtained from the five data sets. We report the across-household means, medians and standard deviations of the household-level posterior means of the purchase rate, recovery rate and drop-out rate, from each of model. We also report the posterior mean of the share of households allocated to the Erlang($k=2$) component.

[Table 6]

Our starting point for the analysis is the result from the EMS model, which we use to compute across-household means of $p$ and $\gamma$. The means of $p$ and $\gamma$ enable us to assess the firm-specific interplay of dropout and recovery, given the purchase behavior of average customers. Across the five data sets, we find significant differences in dropout/recovery behavior for these customers. For the drugstore and the consumer electronics purchase data, we find low across-household means for both $p$ and $\gamma$. In the drugstore data set, the average probability of customers becoming inactive is 4%, in the electronics data set it is 9%. The across-household means of $\gamma$ are 0.11 and 0.15, respectively. This implies that the average inactive drugstore chain customer recovers in 9.2 weeks ($1/0.11$) after dropping out. The average inactive electronics customer recovers after 6.8 weeks ($1/0.15$). For the CDNow and the video equipment data, $p$ is much higher (40%). Thus, customers of these firms drop out more frequently. The average recovery rate of inactive CDNow customers is 0.03 (average downtime: 30.3 weeks) and 0.08 (12.5 weeks) for video equipment customers. The results from the food data are unique, as they combine a low
average probability to become inactive (8%) with a high recovery rate (0.35). From these results, it is apparent that the five data sets describe very different purchase-timing patterns. Food discounter customers rarely become inactive and when they do, they recover fast. Drugstore and consumer electronics customers display a similar propensity to become inactive, but inactive spells are much longer. CDNow and video equipment customers become inactive after almost every second purchase. This event is then followed by long inactive periods.

[Figure 2]

Figure 2 plots the across-household means of $p$ and $\gamma$ over the five data sets. The lower right corner in Figure 2 represents data where the purchasing behavior of average households approaches a BTYD pattern; the probability of defecting is relatively high and the downtime following defection relatively long. CDNow and the video equipment customers reveal this particular type of purchase pattern. The lower left corner in Figure 2 represents data with long downtimes, but a small probability of defecting. The food data, positioned in the upper left corner of Figure 2, combines rare defection events with short downtimes. Such purchase data approach an “always active” pattern. The data sets we have chosen for this analysis span a wide range of activity/inactivity purchase patterns.

Our mixture model also identifies the share of more regular buyers (Table 6). We observe the highest shares of more regular buyers in the video equipment data. 98% of customers of this firm exhibit non-random purchasing behavior. The food data and the drugstore data contain a higher share of randomly buying customers, but regular buying is still dominant (shares of 84% and 79%, respectively for these two data sets). The CDNow data and the consumer electronics data contain mostly randomly purchasing customers. Again, this result is interesting, as it implies that the CDNow data suit the NBD model well.
When comparing results from the different models, it is important to compare the change in the random effects, when moving from the always-active model to BTYD to the state-switching model. A priori, it seems reasonable to expect $\lambda$ to increase as a result of that move, because the (temporary) defection process allows for the interpurchase intervals to be divided into an inactive and an active period. Active periods cannot be longer than the observed interpurchase times, resulting in an increase in $\lambda$. This increase in $\lambda$ should be higher when $p$ increases, because a higher $p$ results in more inactive periods. When moving from the BTYD to the state-switching model, $p$ should increase, because the state-switching model simply provides for more opportunities to defect. The empirical results reported in Table 6 are consistent with these considerations. When allowing for temporary instead of terminal defection, the mean of the dropout probability $p$ increases across all data sets. For four of the five data sets, the mean of the purchase rate $\lambda$ increases when we move from “always active” to state-switching. This increase is largest for the video equipment data, where the mean of $\lambda$ increases from 0.16 to 0.72 when moving from “always active” to state-switching. The increase in the mean of $\lambda$ is also significant in the CDNow data (0.033 to 0.118). The CDNow and the video data yield the largest mean for $p$ across all data sets. In contrast, the increase in $\lambda$, when allowing for state switching, is zero in the electronics data, where $p$ is small.

Overall, we believe that the estimation results have high face validity. Food items are typically purchased at a higher frequency than drugstore items or CDs. Additionally, food shopping trips are often regular events, resulting in non-random purchase intervals. Consequently, the share of non-random buyers who defect rarely and, after defection (i.e. a trip to a different store), recover relatively fast, should be high. In contrast, digital music media are typically purchased less often and much less regularly. Therefore, it is a reasonable result that state-switching occurs more often and that downtimes after defection are longer.
Estimation Results: Household Level

Figure 3 plots the posterior means of $p$ and $\gamma$ on the household level for the food (left) and the CDNow data (right). We have chosen these two data sets of the household-level analysis, as they present polar cases among the data sets analyzed (Figure 2). Figure 3 provides insights into the heterogeneity of unobserved purchasing patterns among customers. In Figure 3, the $p/\gamma$ combinations are visually segmented by the number of purchases. Figure 3 reveals significant heterogeneity in $p$ and $\gamma$ at the same level of activity. Thus, customers with the same number of purchases have quite different timing patterns, yielding different dropout/recovery “stories”. For example, the majority of customers in the CDNow data purchase only once or twice. Yet, the timing of these purchases can be very different. As a result, the range of $\gamma$ in this group is large.

A similar observation can be made for food discounter customers who purchase more than 50 times. The dropout probabilities in this segment span a range from 0 to 0.3, indicating very different dropout behavior.

Specific $p/\gamma$ combinations give rise to different “defection stories”. Describing these patterns across customers is equivalent to segmenting the customer base according to unobserved drivers of purchasing behavior. A high $p$/small $\gamma$ combination is equivalent to purchasing until terminal defection (BTYD). The CDNow data set contains many of these customers, so that we find the EMS to perform relatively poorly, compared to the BTYD model in this data set. However, we can observe some customers in this data set who recover fast, contradicting the BTYD assumption. In the food data set, customers span a much wider range of $p$ and $\gamma$, implying more heterogeneity and greater diversity in unobserved purchase timing patterns. Therefore, it is more difficult to distinguish between purchase-timing segments that are conditional on observable purchasing activity. There is more overlap in the distributions of $p$ and
\(\gamma\) at the same activity levels. Thus, the observed number of transactions is not very informative about the unobservable state-switching process.

6. Discussion

In this paper, we propose an Erlang-mixture state-switching model which makes two contributions to the literature: (1) The proposed EMS model enables us to distinguish between the activity and inactivity of customers. In doing so, it captures the heterogeneity among customers with regard to the interplay of temporary drop-out and recovery. This heterogeneity can be used to segment the customer base according to unobservable aspects of purchase timing. The proposed EMS model identifies customers who are “always active”, transient switchers between an inactive and an active state, and BTYD-type buyers. Given individual-level estimates of \(p/\gamma\), we can allocate customers to state-based groups. Given customer cohort data, it is possible to analyze changes in these segments over time. Such an analysis would provide insights such as: Does the share of BTYD buyers increase over time in my customer base? Is the average recovery rate decreasing or increasing across cohorts? These are interesting and useful indicators of customer equity dynamics. (2) Our model captures heterogeneity among households with regard to the regularity of buying. Across the data sets, we find that a “one size fits all” approach (all households purchase randomly or more regularly) is not realistic. To the best of our knowledge, no other model captures this type of heterogeneity. Our results indicate that it is important to model buying as a mixture of random and more regular events. The share of non-random buyers varies greatly across the data sets analyzed and approaches 100% in one particular case. Under such circumstances, it is necessary to relax the assumption of random buying across all households.
The application of our model to multiple data sets demonstrates that the proposed EMS model is generally preferred to those assuming terminal defection or no defection. The implications of this result are significant, given that many benchmark models in customer scoring are BTYD models. Of course, when customer behavior is predominantly BTYD, our model provides little advantage over others. However, the question of whether BTYD is appropriate or not is an empirical issue, and our model treats it as such.

Our results present a number of future research opportunities. It would be interesting to externally validate unobserved switching between the inactive and the active state. Also, it would be worth using covariates for the recovery rate, purchase rate and drop-out probability, in order to make inferences about the drivers of these latent variables. Covariates such as household size and purchase volume might explain why households recover at a certain rate from inactivity or become inactive. It would also be interesting to pursue the question of whether $p/\gamma$ combinations are indicators of share-of-wallet. Given a certain household-size (or size of wallet), a high $p$ and small $\gamma$ might indicate a low share-of-wallet, compared to a household with a small $p$ and high $\gamma$, complementing existing approaches for making inferences about customers’ unobserved share of wallet (Du et al. 2007). Extensions of the proposed model include allowing for a priori correlated random effects and multiple Erlang mixture components of a higher order. The latter extension might be promising, for instance, for food discounter data, for which some buyers exhibit a clockwise purchase pattern. For this type of customer, fixing $k$ to a relatively small value (e.g. $1<k<10$) might not be appropriate.
References


Appendix A: Prior Specification and MCMC Sampling Procedure

Prior Set-up

We employ the following weakly-informative subjective priors in our analysis:

(1) For the membership to the Erlang components:

\[ p(\eta) = \text{Beta}(3,3) \]

(2) We employ a mixture of prior distributions for the household-level quantities:

\[ p(\lambda) \sim G(a_{\lambda,i}, b_{\lambda,i}), l_i \sim \text{Multinomial}_i(\omega_\lambda) \]

where \( \omega \) indicates the size of the prior components. Similarly:

\[ p(\gamma) \sim G(a_{\gamma,j}, b_{\gamma,j}), l'_i \sim \text{Multinomial}_i(\omega_\gamma) \]

\[ p(p) \sim B(a_{p,i}, b_{p,i}), l''_i \sim \text{Multinomial}_i(\omega_p) \]

with \( a=1, b=0.1 \) for all Gamma distributions and \( a=3, b=3 \) for the Beta distributions. The hyperprior for the prior component membership is:

\[ p(\omega) = \text{Dirichlet}(3,..,3) \]

MCMC Estimation Procedure

Step 1: Augmentation of State Sequence and Recovery Waiting Times

Estimation of the EMS relies upon the augmentation of the unobserved sequence of states that customers go through (active, inactive) and on obtaining estimates of the unobserved recovery waiting times (see Figure 1 for a directed acyclic graph of our model). For convenience, we set \( t_0 = 0 \) and \( t_{n+1} = T \) where \( n \) denotes the number of transactions a certain customer made within \( T \). 0 and \( T \) denote the start and the end of the observation period, respectively. \( t_j \) is the time of the \( j \)-th transaction. To reduce notational clutter, we suppress the customer index \( i \) and
the mixture index $q$ in the following. The observed waiting times between transactions at $t_{j-1}$ and $t_j$ are denoted $\Delta t_j$. The waiting times can be arranged in a vector of length $n+1$ with the last element being the time between the last transaction and $T$ ($\Delta t_{n+1}$). For all intervals $\Delta t_j$ from $t_0$ up to $t_n$, let $s_j=S$ denote the state in which customer are, taking on values:

- $S_1$: active throughout $\Delta t_j$
- $S_2$: inactive (dropout at $t_{j-1}$ and recovery before $t_j$)

For the interval $\Delta t_{n+1}$, let $S'$ denote the state in which customers are:

- $S_1'$: active throughout $\Delta t_{n+1}$
- $S_2'$: inactive throughout $\Delta t_{n+1}$ (recovery after $T$)
- $S_3'$: inactive and recovery before $T$

[Figure 1]

Becoming inactive during any interval up to the last purchase event implies that a customer recovers during all of them since we observe another purchase. In the last interval, this is obviously different. A customer can stay inactive since we do not observe a purchase at $T$ (we call this state $S_2'$). She may also recover ($S_3'$), but then the next purchase must occur after $T$.

The DAG of our model (Figure 1) shows that, conditional on the state, $\Delta t_j$ is a mix of recovery times and transactional waiting times. If a customer becomes inactive, the time until recovery is $w_j$ and this waiting time is determined by the rate of the recovery process, $\gamma$. Becoming inactive and to recover before $t_j$ implies that the transactional waiting time (waiting time after recovery until next purchase) is $\Delta t_j - w_j$. If the customer stays active, $w_j=0$. These considerations imply that, given $\theta = (\lambda, \gamma, p, k)$ and $\Delta t$, we can factor the joint distribution of the state and recovery times as follows:
Using (1), we first obtain draws of the state sequence and then, conditional on the state, draws of the recovery times. Note that given \( s_j \) and \( \theta = (\lambda, \gamma, p, k) \), the distribution of \( \Delta t_j \) is independent of events outside interval \( j \). Thus, \( p(\Delta t_j | S_j, \{ \Delta t_1, \Delta t_2, ..., \Delta t_{j-1} \}, \theta) = p(\Delta t_j | S_j, \theta) \).

Given \( S \) and \( \theta \), \( \begin{bmatrix} \Delta t_j | S_1, \theta \end{bmatrix} \sim \text{Erlang}(k, \lambda) \) and \( \begin{bmatrix} \Delta t_j | S_2, \theta \end{bmatrix} \sim \text{Exp}(\gamma) \times \text{Erlang}(k, \lambda) \). The likelihood of observing \( \Delta t_j \) given \( S_2 \) is then:

\[
p(\Delta t_j | S_j = S_2, \theta) = \int_0^{\Delta t_j} y e^{-y} \frac{\lambda^k (\Delta t_j - y)^{k-1} e^{-\lambda(\Delta t_j - y)}}{(k-1)!} \, dy
\]

\[
= \begin{cases} 
\gamma e^{-\gamma \Delta t_j} \left( \frac{\lambda}{\lambda - \gamma} \right)^k \Gamma(k,(\lambda - \gamma)\Delta t_j), & \lambda \neq \gamma \\
\lambda^{k+1} \Delta t_j^k e^{-\lambda \Delta t_j} / \Gamma(k+1), & \lambda = \gamma
\end{cases}
\]

where \( \Gamma(k,x) = 1 - e^{-x} \sum_{m=0}^{k-1} \frac{x^m}{m!} \), and \( \Gamma(k+1) = k! \)

The posterior distribution of \( s_j \) is:

\[
p(s_j = S_1 | \theta, \Delta t_j) = \frac{L(\Delta t_j | \theta, s_j = S_1) p(s_j = S_1 | \theta)}{p(\Delta t_j | \theta)}
\]

\[
= \begin{cases} 
\frac{1}{1 + p \gamma e^{(\lambda-\gamma)\Delta t_j} \Gamma(k,(\lambda - \gamma)\Delta t_j)} , & \lambda \neq \gamma \\
\frac{1}{1 + p \lambda \Delta t_j / k(1 - p)} , & \lambda = \gamma
\end{cases}
\]

\[
p(s_j = S_2 | \theta, \Delta t_j) = 1 - p(s_j = S_1 | \theta, \Delta t_j)
\]
Expressions (4) and (5) provide the probabilities to draw a state sequence for all interpurchase intervals up to \( n \), using a binomial distribution. In the last interval: \( j = n+1 \) and \( \Delta t_{n+1} = T - t_n \). The probability that the customer is in state \( S_1' \) is:

\[
P(s_{n+1} = S_1', \Delta t_{n+1} | \theta) = \left(1 - p\right) \sum_{m=0}^{k-1} e^{-\lambda \Delta t_{n+1}} \left( \lambda \Delta t_{n+1} \right)^m / m! = \left(1 - p\right) \left[1 - \Gamma(k, \lambda \Delta t_{n+1})\right].
\]  

(6)

The probability of the customer being in state \( S_2' \) is:

\[
P(s_{n+1} = S_2', \Delta t_{n+1} | \theta) = pe^{-\gamma \Delta t_{n+1}}.
\]  

(7)

The probability that the customer is in state \( S_3' \) is:

\[
P(s_{n+1} = S_3', \Delta t_{n+1} | \theta) = \int_0^{\Delta t_{n+1}} pye^{-\gamma t} \sum_{m=0}^{k-1} e^{-\lambda (\Delta t_{n+1} - t)} \left( \lambda (\Delta t_{n+1} - t) \right)^m / m! dt
\]

\[
= \begin{cases} 
pe^{-\gamma \Delta t_{n+1}} & \lambda = \gamma \\
\sum_{m=0}^{k-1} \frac{\lambda^m}{(\lambda - \gamma)^{m+1}} \Gamma(m + 1, (\lambda - \gamma) \Delta t_{n+1}), \lambda \neq \gamma & \lambda \neq \gamma
\end{cases}
\]  

(8)

Then the posterior probability of \( s_{n+1} \) is:

\[
P(s_{n+1} = S_i' | \theta, \Delta t_{n+1}) = \frac{P(s_{n+1} = S_i', \Delta t_{n+1} | \theta)}{\sum_i P(s_{n+1} = S_i', \Delta t_{n+1} | \theta)}, i = 1, 2, 3.
\]  

(9)

Using expressions (7)-(9), we can draw the last state via a multinomial distribution. The conditional independence properties satisfied by \( w_j \) allow it to be drawn given \( s_j \). The conditional distribution of \( w_j \) is:
\[ p(w_j | s_j, \Delta t_j, \theta) = \begin{cases} 
0 & \text{with probability 1 if } s_j = S_1, S'_1 \\
\Delta t_j & \text{with probability 1 if } s_j = S'_2 \\
\frac{(-\gamma)^k (\Delta t_j - w_j)^{k-1} e^{-(\lambda - \gamma) (\Delta t_j - w_j)}}{\Gamma(k) \Gamma(k, (\lambda - \gamma) \Delta t_j)} & \text{if } s_j = S_2, S'_2, \lambda \neq \gamma \\
\frac{k}{\Delta t_j} \left( 1 - \frac{w_j}{\Delta t_j} \right)^{k-1} & \text{if } s_j = S_2, S'_2, \lambda = \gamma 
\end{cases} \quad (10) \]

\( w_j \in (0, \Delta t_j), j = 1, \ldots, n+1 \) and \( \Gamma(k) = (k-1)! \). When \( \lambda \geq \gamma \), the distribution \( p(w_j|s_j=S_2 \text{ or } S'_3, \Delta t_j, \theta) \) can be sampled via CDF inversion:

\[ F^{-1}(\xi) = \begin{cases} 
\frac{1}{\lambda - \gamma} \Gamma^{-1}(k, \Gamma(k, (\lambda - \gamma) \Delta t_j) \xi), \lambda > \gamma \\
\Delta t_j - (1 - \xi)^k, \lambda = \gamma 
\end{cases} \]

where \( \Gamma^{-1}(k, x) \) is an inverse incomplete gamma function and \( \xi \) is uniformly distributed on \([0,1]\). When \( \lambda < \gamma \), we use slice-sampling (Neil 2003) to obtain random samples of \( w_j \). After drawing the state sequence \( S \) and the recovery time sequence \( w \), we can collect the information contained in these draws:

\[ n_{01} = \sum_{j=1}^{n+1} I(s_j = S_2 \text{ or } S'_3), \quad n_{10} = \sum_{j=1}^{n+1} I(s_j = S_2, S'_2 \text{ or } S'_3) \]

\[ T_0 = \sum_j (w_j), \quad T_1 = \sum_j (\Delta t_j - w_j) = T - T_0 \]

\[ T_p = \prod_{j=1}^n (\Delta t_j - w_j)^{I(\Delta t_j > w_j)}, \quad t_p = \Delta t_{n+1} - w_{n+1} \]

where \( n_{01} \) is the number of times a customer recovers from the inactive state, \( n_{10} \) is the number of switches to the inactive state, \( T_0 \) is the total time spent in the inactive state, \( T_1 \) is the total time
spent in the active state up to transaction $n$, $T_p$ is the product of the active periods up to transaction $n$ and $t_p$ is the length of the active period in the last interval.

**Step 2: Estimation of Household-Level Quantities**

Draws of $\gamma$ and $p$ are obtained using the information collected across purchase intervals and conjugate results:

$$p(\gamma | \cdot) \propto p(w | \gamma, S) \times p(\gamma | a_\gamma, b_\gamma) = p(T_0, n_0 | \gamma) \times p(\gamma | a_\gamma, b_\gamma) = \text{Gamma}(n_0 + a_\gamma, T_0 + b_\gamma)$$

$$p(p | \cdot) \propto p(S | p) \times p(p | a_p, b_p) = p(n_{10}, n_0 | n, p) \times p(p | a_p, b_p) = \text{Beta}(n_{10} + a_\gamma, n - n_0 + b_\gamma)$$

The conditional posterior distribution of $\lambda$ is given by:

$$p(\lambda | T_1, n, a_\lambda, b_\lambda) \propto p(T_1, n | \lambda, k, a_\lambda, b_\lambda) \times p(\lambda | a_\lambda, b_\lambda) =$$

$$p\left(\left\{ (\Delta t_1 - w_1), (\Delta t_2 - w_2), ..., (\Delta t_{n+1} - w_{n+1}) \right\} \big| \lambda, k, a_\lambda, b_\lambda \right) \times p(\lambda | a_\lambda, b_\lambda) =$$

$$\prod_{j=1}^{n} \frac{\lambda^k (\Delta t_j - w_j)^{k-1}}{(k-1)!} e^{-\lambda(\Delta t_j - w_j)} \times \left( e^{-\lambda(\Delta t_{n+1} - w_{n+1})} \sum_{m=0}^{k-1} \frac{(\lambda(\Delta t_{n+1} - w_{n+1}))^m}{m!} \right) \times \frac{b_\lambda^{a_\lambda}}{\Gamma(a_\lambda)} \lambda^{a_\lambda - 1} e^{-\lambda b_\lambda}$$

where the second term in the third line gives the probability of not observing a purchase between recovery and $T$. The above expression can be written as:

$$p(\lambda | T_1, n, k, a_\lambda, b_\lambda) \propto \lambda^{k+a_\lambda-1} e^{-\lambda(T_1+b_\lambda)} \times u \quad (11)$$

where:

$$u = \begin{cases} 
\sum_{m=0}^{k-1} \frac{(\lambda(\Delta t_{n+1} - w_{n+1}))^m}{m!} & \text{if } S_{n+1} = S_1 \text{ or } S_{n+1} = S_3 \\
1 & \text{if } S_{n+1} = S_2 
\end{cases}$$

Thus, if customers are inactive throughout $\Delta t_{n+1}$, the draw of $\lambda$ is a conjugate draw from a gamma density. In the other cases, a conjugate draw is not available and we draw $\lambda$ using a slice sampler.
The update of the parameters of the prior distributions of \( \lambda \), \( \gamma \), and \( p \) is identical to Ma and Büschken (2011). To reduce the influence of the hyperpriors for households with few purchases and therefore scarce information on the likelihood level, we model a mixture of hyperprior distributions.

**Step 3: Erlang Component Membership**

The Erlang mixture model is implemented via the latent component membership indicator \( k_q \) that we update at every step of the MCMC chain (Carlin and Chib 1995, Frühwirth-Schnatter 2006). The posterior distribution of \( k_q \) is given by:

\[
p(k_q | \Delta t, \eta, \lambda, Q) = \frac{p(T_i | k_q, \lambda) \times p(k_q | \eta_q)}{\sum_{q=1}^{Q} p(T_i | k_q, \lambda) \times p(k_q | \eta_q)}
\]  \hspace{1cm} (12)

(12) provides the posterior probability of \( k_q \), given accumulated time in the active state. The weight of component \( q \) (\( \eta_q \)) is updated via a Dirichlet draw.
Appendix B: Key Managerial Expressions

In the following, we develop two key managerial expression for the ESM: $E(X(T))$ and $E(Y(t))$. The first quantity refers to the expected number of transactions of a household given $T$ and individual-level parameters $\theta = (\lambda, \gamma, p, k)$, the second quantity refers to the expected number of transactions in the holdout period given $\theta$. For $k=1$, we use the expressions derived in Ma and Büschken (2011). Here, we here demonstrate how equivalent expressions for $k=2$ can be derived.

Interpurchase times follow an Erlang(2, $\lambda$) distribution which is the sum of two independent exponential distributions, each with parameter $\lambda$. This corresponds to recording every second event as a “real transaction”, preceded by a “dummy transaction” which we don’t record. At any time, a customer may stay in one of three states: Active1, Active2, and Inactive. Active1 is an active state before a dummy purchase is made; Active2 represents an active state after a dummy transaction and before a real purchase is made. “Inactive” represents the dormant state which is only accessible after a real purchase. To derive our key managerial expressions, we uniformize our model using the uniformization rate $\omega = \lambda + \gamma$. Given this uniformization, the purchasing process is equivalent to a Markov process that spends an $i.i.d.$ exponential ($\omega$) amount of time in each state and has the following transition probabilities among states (Ho et al. 2005):

$$
P = \begin{pmatrix}
\frac{\gamma}{\omega} & \frac{\lambda}{\omega} & 0 \\
(1-p)\frac{\lambda}{\omega} & \frac{\gamma}{\omega} & p\frac{\lambda}{\omega} \\
\frac{\gamma}{\omega} & 0 & \frac{\lambda}{\omega}
\end{pmatrix}
$$

Suppose the customer starts in state $i$ and has had $n$ transitions in the uniformized Markov chain. Because each real transition from state 2 to other states corresponds to a purchase by the customer, we would like to know how many of the $n$ arrivals correspond to real purchases. Let $N_i(n)$ be the expected number of purchases in the next $n$ transitions of the uniformized embedded Markov chain, where $i \in \{1, 2, 3\}$. We then have the following:

$$
N_1(n) = \frac{\gamma}{\omega} N_1(n-1) + \frac{\lambda}{\omega} N_2(n-1),
$$

$$
N_2(n) = \frac{(1-p)\lambda}{\omega} (1 + N_1(n-1)) + \frac{p\lambda}{\omega} (1 + N_3(n-1)) + \frac{\gamma}{\omega} N_2(n-1)
= \frac{\lambda}{\omega} + \frac{(1-p)\lambda}{\omega} N_1(n-1) + \frac{p\lambda}{\omega} N_3(n-1) + \frac{\gamma}{\omega} N_2(n-1)
$$
\[ N_3(n) = \frac{\gamma}{\omega} N_3(n-1) + \frac{\lambda}{\omega} N_3(n-1) \].

If we let:

\[
N(n) = \begin{pmatrix} N_1(n) \\ N_2(n) \\ N_3(n) \end{pmatrix}, \quad Z = \begin{pmatrix} 0 \\ \frac{\lambda}{\omega} \\ 0 \end{pmatrix}
\]

Then:

\[
N(n) = Z + PN(n-1) = \left( \sum_{k=0}^{n-1} P^k \right) Z \tag{A.1}
\]

Let \( E(X(t)) \) be the expected number of transactions in a time period of length \( t \), which is central to the computation of the expected transaction volume for the whole customer base over time. We assume that at time 0 a customer makes the initial purchase. At that time, our belief state of the customer is \( B(0) = \begin{pmatrix} 1-p \\ 0 \\ p \end{pmatrix} \). This is because, after the initial purchase, a customer cannot switch immediately to state \( A2 \). It follows that:

\[
E(X(t)) = \sum_{n=0}^{\infty} B(0)^T \cdot N(n) \frac{e^{-\omega t} (\omega t)^n}{n!} \tag{A.2}
\]

A.1 and A.2 can be combined to compute \( E(X(t)) \). A closed-form expression for \( E(X(t)) \) as in the BG-NBD model or the HIPP model is not available. However, we can calculate the above quantity at the desired level of accuracy by using a sufficiently large upper bound for the sums on the RHS of the equation above. Let the random variable \( Y(t) \) denotes the number of purchases made during the holdout period \( (T, T+t] \). In order to calculate \( E(Y(t)) \), we need to estimate the belief state of the customer at time \( T \). There are five possible situations in the last interval \( \Delta t_{n+1} \) which are illustrated in the figure below. The circles in the states \( St_1 \) and \( St_4 \) represent a dummy purchase.

![Diagram of belief states](image-url)
St₁ represents the state in which a dummy transaction has been recorded, but a real purchase has not been made, yet. St₂ is the state after a real transaction and before a dummy transaction. This implies that the customer did not become inactive after the real transaction. St₃ is the inactive state, after defection and before the recovery has occurred. St₄ represents a state after recovery and after the switch to A2 and, thus, before a real purchase. St₅ is after recovery and before the dummy transaction. The likelihood for each state can be written down as follows:

\[ P(St_1) = (1 - p) \int_0^{\Delta t_{n+1}} \lambda e^{-\lambda \tau} e^{-\lambda (\Delta t_{n+1} - \tau)} d\tau = (1 - p) \lambda \Delta t_{n+1} e^{-\lambda \Delta t_{n+1}}, \]

\[ P(St_2) = (1 - p) e^{-\lambda \Delta t_{n+1}}, \quad P(St_3) = pe^{-\gamma \Delta t_{n+1}}, \]

\[ P(St_4) = p \int_0^{\Delta t_{n+1}} \gamma e^{-\gamma t} \int_0^{\Delta t_{n+1} - t} \lambda e^{-\lambda \tau} e^{-\lambda (\Delta t_{n+1} - \tau - t)} d\tau d\tau \]

\[ = \frac{p \gamma \lambda}{(\lambda - \gamma)^2} \left( e^{-\gamma \Delta t_{n+1}} - e^{-\lambda \Delta t_{n+1}} \right) - \frac{p \gamma \lambda}{\lambda - \gamma} \Delta t_{n+1} e^{-\lambda \Delta t_{n+1}} \]

When \( \lambda = \gamma \),

\[ P(St_4) = \frac{1}{2} p \gamma \Delta t^2 e^{-\lambda t}. \]

\[ P(St_5) = p \int_0^{\Delta t_{n+1}} \gamma e^{-\gamma t} e^{-\lambda (\Delta t_{n+1} - t)} d\tau = \frac{p \gamma}{(\lambda - \gamma)} \left( e^{-\gamma \Delta t_{n+1}} - e^{-\lambda \Delta t_{n+1}} \right), \]

When \( \lambda = \gamma \),

\[ P(St_5) = \frac{1}{2} p \gamma \Delta t e^{-\lambda t}. \]

Given individual-level quantities, we can now derive the belief state at T:

\[ B(T) = \left( P(St_2) + P(St_5), P(St_1) + P(St_4), P(St_3) \right) / \sum_i P(St_i). \]

It follows that:

\[ E(Y(t)) = \sum_{n=0}^{\infty} B(T)^T \cdot N(n) \frac{e^{-\alpha t} (\alpha t)^n}{n!} \]
Table 1: Model Overview

<table>
<thead>
<tr>
<th>Always active (p=0 or γ&gt;∞)</th>
<th>Erlang(k=1)</th>
<th>Erlang(k=2)</th>
<th>Mixture (k ∈ {1, 2})</th>
</tr>
</thead>
<tbody>
<tr>
<td>NBD</td>
<td>Condensed NBD</td>
<td>Mixture of Erlang IPT</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Buy-until-you-die (p≥0, γ=0)</th>
<th>BG NBD</th>
<th>BG Condensed NBD</th>
<th>EMS with absorbing inactive state</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>State switching (p≥0, γ≥0)</th>
<th>HIPP</th>
<th>Condensed HIPP</th>
<th>EMS</th>
</tr>
</thead>
</table>

Table 2: Simulation Results

<table>
<thead>
<tr>
<th>Parameters</th>
<th>True Value</th>
<th>Posterior Mean</th>
<th>Posterior SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component Share</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(k=1)</td>
<td>0.5</td>
<td>0.469</td>
<td>0.004</td>
</tr>
<tr>
<td>(k=2)</td>
<td>0.5</td>
<td>0.531</td>
<td>0.004</td>
</tr>
</tbody>
</table>

| Parameters of Heterogeneity Distributions | | | |
| a_{λ} | 10 | 10.55 | 0.552 |
| b_{λ} | 10 | 10.54 | 0.560 |
| a_{γ} | 10 | 9.78 | 0.334 |
| b_{γ} | 100 | 99.7 | 3.017 |
| a_p | 12 | 11.87 | 0.566 |
| a_p | 100 | 99.25 | 3.56 |

| MSE Random Effects | | | |
| \(λ\) | 0.0116 | | |
| \(γ\) | 0.0006 | | |
| \(p\) | 0.0007 | | |

<p>| Hit Rate for Erlang Mixture Components | 0.996 |
| MAPE Accumulated Incative Time | 6.905 |</p>
<table>
<thead>
<tr>
<th></th>
<th>Food</th>
<th>Drugstore Chain</th>
<th>Video Tech</th>
<th>Electronics Retailer</th>
<th>CDNow</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Discounter</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of customers/households</td>
<td>1,638</td>
<td>902</td>
<td>401</td>
<td>2,440</td>
<td>2121</td>
</tr>
<tr>
<td>(holdout sample)</td>
<td>(183)</td>
<td>(101)</td>
<td>(60)</td>
<td>(272)</td>
<td>(236)</td>
</tr>
<tr>
<td>Length of observation period</td>
<td>80*</td>
<td>80*</td>
<td>42*</td>
<td>104*</td>
<td>32.71*</td>
</tr>
<tr>
<td>in weeks (holdout period in weeks)</td>
<td>(24.75)*</td>
<td>(24.75)*</td>
<td>(9) +</td>
<td>(43) +</td>
<td>(39) +</td>
</tr>
<tr>
<td>Number of transactions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>30</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Mean</td>
<td>41.6</td>
<td>12.65</td>
<td>3.13</td>
<td>1.7</td>
<td>1.04</td>
</tr>
<tr>
<td>SD</td>
<td>40.24</td>
<td>15.87</td>
<td>4.98</td>
<td>2.78</td>
<td>2.19</td>
</tr>
<tr>
<td>Interpurchase times (weeks)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>1</td>
<td>2.57</td>
<td>3</td>
<td>18.43</td>
<td>10.14</td>
</tr>
<tr>
<td>Mean</td>
<td>1.87</td>
<td>5.86</td>
<td>6.93</td>
<td>32.75</td>
<td>16.12</td>
</tr>
<tr>
<td>SD</td>
<td>4.27</td>
<td>10.36</td>
<td>9.57</td>
<td>40.26</td>
<td>14.28</td>
</tr>
<tr>
<td>Recency (weeks)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>1.71</td>
<td>6.28</td>
<td>13</td>
<td>31.57</td>
<td>30.57</td>
</tr>
<tr>
<td>Mean</td>
<td>7.72</td>
<td>16.36</td>
<td>16.23</td>
<td>52.26</td>
<td>26.30</td>
</tr>
<tr>
<td>SD</td>
<td>14.38</td>
<td>20.47</td>
<td>14.04</td>
<td>55.68</td>
<td>11.15</td>
</tr>
<tr>
<td>Average recency/average</td>
<td>0.097</td>
<td>0.205</td>
<td>0.386</td>
<td>0.503</td>
<td>0.804</td>
</tr>
<tr>
<td>observation period</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*: homogeneous across hh (panel data); +: average value across hh
<table>
<thead>
<tr>
<th></th>
<th>Food Discounter</th>
<th>Drugstore Chain</th>
<th>Video Equipment</th>
<th>Electronics Retailer</th>
<th>CDNow</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Always Active</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-sample Fit</td>
<td>-76,064.80</td>
<td>-25,068.03</td>
<td>-3,176.17</td>
<td>-19,610.00</td>
<td>-8,009.18</td>
</tr>
<tr>
<td>Predictive Fit</td>
<td>-23,318.16</td>
<td>-6,596.19</td>
<td>-252.62</td>
<td>-1,664.31</td>
<td>-1,577.13</td>
</tr>
<tr>
<td><strong>Buy-until-you-die</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-sample Fit</td>
<td>-75,841.90</td>
<td>-25,021.66</td>
<td>-3,203.19</td>
<td>-19,648.26</td>
<td>-7,951.53</td>
</tr>
<tr>
<td>Predictive Fit</td>
<td>-22,716.74</td>
<td>-6,587.02</td>
<td>-246.28</td>
<td>-1,705.38</td>
<td>-1,302.03</td>
</tr>
<tr>
<td><strong>State Switching</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-sample Fit</td>
<td>-74,512.20</td>
<td>-24,953.83</td>
<td>-3,145.57</td>
<td>-19,607.78</td>
<td>-7,835.54</td>
</tr>
<tr>
<td>Predictive Fit</td>
<td>-20,173.08</td>
<td>-6,417.90</td>
<td>-193.61</td>
<td>-964.81</td>
<td>-1,467.14</td>
</tr>
</tbody>
</table>
Table 5: Household-Level Predictions: MSE of $E(Y(t)|\theta_i)$

<table>
<thead>
<tr>
<th></th>
<th>Food Discounter</th>
<th>Drugstore</th>
<th>Electronics Retailer</th>
<th>CDNow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve Prediction</td>
<td>27.51</td>
<td>7.85</td>
<td>2.21</td>
<td>4.85</td>
</tr>
<tr>
<td>Always Active</td>
<td>28.23</td>
<td>7.67</td>
<td>1.06</td>
<td>3.52</td>
</tr>
<tr>
<td>Buy-until-you-die</td>
<td>28.76</td>
<td>7.29</td>
<td>0.99</td>
<td>2.57</td>
</tr>
<tr>
<td>State Switching</td>
<td>26.87</td>
<td>7.17</td>
<td>1.04</td>
<td>2.56</td>
</tr>
</tbody>
</table>
Table 6: Estimation Results

<table>
<thead>
<tr>
<th>Drugstore</th>
<th>Median</th>
<th>Mean</th>
<th>SD</th>
<th>Food Retailer</th>
<th>Median</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>State switching</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gamma*</td>
<td>0.089</td>
<td>0.109</td>
<td>0.05</td>
<td></td>
<td>0.356</td>
<td>0.349</td>
<td>0.087</td>
</tr>
<tr>
<td>lambda</td>
<td>0.172</td>
<td>0.317</td>
<td>0.375</td>
<td></td>
<td>0.913</td>
<td>1.153</td>
<td>1.036</td>
</tr>
<tr>
<td>p</td>
<td>0.044</td>
<td>0.044</td>
<td>0.028</td>
<td></td>
<td>0.079</td>
<td>0.080</td>
<td>0.054</td>
</tr>
<tr>
<td>k=2 (share of)</td>
<td>0.789</td>
<td>-</td>
<td>-</td>
<td></td>
<td>k=2 (share of)</td>
<td>0.837</td>
<td></td>
</tr>
<tr>
<td><strong>Buy Until You Die</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gamma</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>lambda</td>
<td>0.150</td>
<td>0.294</td>
<td>0.354</td>
<td></td>
<td>0.665</td>
<td>0.973</td>
<td>0.956</td>
</tr>
<tr>
<td>p</td>
<td>0.005</td>
<td>0.016</td>
<td>0.031</td>
<td></td>
<td>0.0005</td>
<td>0.003</td>
<td>0.011</td>
</tr>
<tr>
<td>k=2 (share of)</td>
<td>0.733</td>
<td>-</td>
<td>-</td>
<td></td>
<td>k=2 (share of)</td>
<td>0.741</td>
<td></td>
</tr>
<tr>
<td><strong>Always Active</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gamma</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>lambda</td>
<td>0.130</td>
<td>0.277</td>
<td>0.356</td>
<td></td>
<td>0.647</td>
<td>0.968</td>
<td>0.961</td>
</tr>
<tr>
<td>p</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>k=2 (share of)</td>
<td>0.682</td>
<td>-</td>
<td>-</td>
<td></td>
<td>k=2 (share of)</td>
<td>0.726</td>
<td></td>
</tr>
</tbody>
</table>

* cross-sectional means of household-level posterior means

<table>
<thead>
<tr>
<th>CDNow</th>
<th>Median</th>
<th>Mean</th>
<th>SD</th>
<th>Video Equipment</th>
<th>Median</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>State switching</strong></td>
<td></td>
<td></td>
<td></td>
<td><strong>State switching</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gamma</td>
<td>0.025</td>
<td>0.033</td>
<td>0.022</td>
<td></td>
<td>0.036</td>
<td>0.076</td>
<td>0.081</td>
</tr>
<tr>
<td>lambda</td>
<td>0.076</td>
<td>0.118</td>
<td>0.114</td>
<td></td>
<td>0.702</td>
<td>0.723</td>
<td>0.138</td>
</tr>
<tr>
<td>p</td>
<td>0.450</td>
<td>0.424</td>
<td>0.055</td>
<td></td>
<td>0.474</td>
<td>0.450</td>
<td>0.041</td>
</tr>
<tr>
<td>k=2 (share of)</td>
<td>0.190</td>
<td>-</td>
<td>-</td>
<td></td>
<td>k=2 (share of)</td>
<td>0.976</td>
<td></td>
</tr>
<tr>
<td><strong>Buy Until You Die</strong></td>
<td></td>
<td></td>
<td></td>
<td><strong>Buy Until You Die</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gamma</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>lambda</td>
<td>0.078</td>
<td>0.103</td>
<td>0.091</td>
<td></td>
<td>0.440</td>
<td>0.454</td>
<td>0.151</td>
</tr>
<tr>
<td>p</td>
<td>0.453</td>
<td>0.389</td>
<td>0.116</td>
<td></td>
<td>0.231</td>
<td>0.201</td>
<td>0.069</td>
</tr>
<tr>
<td>k=2 (share of)</td>
<td>0.139</td>
<td>-</td>
<td>-</td>
<td></td>
<td>k=2 (share of)</td>
<td>0.970</td>
<td></td>
</tr>
<tr>
<td><strong>Always Active</strong></td>
<td></td>
<td></td>
<td></td>
<td><strong>Always Active</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gamma</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>lambda</td>
<td>0.010</td>
<td>0.033</td>
<td>0.068</td>
<td></td>
<td>0.093</td>
<td>0.156</td>
<td>0.170</td>
</tr>
<tr>
<td>p</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>k=2 (share of)</td>
<td>0.032</td>
<td>-</td>
<td>-</td>
<td></td>
<td>k=2 (share of)</td>
<td>0.461</td>
<td></td>
</tr>
</tbody>
</table>

39
Table 6: Estimation Results (cont’d)

<table>
<thead>
<tr>
<th>Electronics Retailer</th>
<th>Median</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>State switching</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gamma</td>
<td>0.147</td>
<td>0.147</td>
<td>0.004</td>
</tr>
<tr>
<td>lambda</td>
<td>0.015</td>
<td>0.019</td>
<td>0.016</td>
</tr>
<tr>
<td>p</td>
<td>0.087</td>
<td>0.086</td>
<td>0.017</td>
</tr>
<tr>
<td>k=2 (share of)</td>
<td>0.054</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Buy Until You Die</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gamma</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>lambda</td>
<td>0.019</td>
<td>0.023</td>
<td>0.016</td>
</tr>
<tr>
<td>p</td>
<td>0.076</td>
<td>0.076</td>
<td>0.034</td>
</tr>
<tr>
<td>k=2 (share of)</td>
<td>0.129</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Always Active</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gamma</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>lambda</td>
<td>0.015</td>
<td>0.019</td>
<td>0.016</td>
</tr>
<tr>
<td>p</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>k=2 (share of)</td>
<td>0.064</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1: DAG of the state switching model (without mixture of prior distributions for $p, \lambda, \gamma$)
Figure 2: Across-household means of $p$ and $\gamma$ given data set
Figure 3: Household-level posterior means of $p$ (dropout probability) and $\gamma$ (recovery rate) for food discounter data (above) and CDNow data (below) \(^3\)

---

3 The legend indicates the observed number of transactions (e.g.: “>100” identifies food-discounter customer (upper graph), who made more than 100 transactions in the calibration period). Some CDNow customers (lower graph) make no repeat purchase (zero transactions, indicated by the black segment). For these customers, $p$ and $\gamma$ are drawn by using only the prior.