# "Counting Your Customers": When will they buy next? An empirical validation of probabilistic customer base analysis models based on purchase timing

E. Korkmaz R. Kuik D. Fok

08–01–2013

**Abstract**

This research provides a new way to validate and compare buy-till-you-defect [BTYD] models. These models specify a customers transaction and defection processes in a non-contractual setting. They are typically used to identify active customers in a com- panys customer base and to predict the number of purchases. Surprisingly, the literature shows that models with quite different assumptions tend to have a similar predictive performance. We show that BTYD models can also be used to predict the timing of the next purchase. Such predictions are managerially relevant as they enable managers to choose appropriate promotion strategies to improve revenues. Moreover, the predictive performance on the purchase timing can be more informative on the relative quality of BTYD models. For each of the established models, we discuss the prediction of the purchase timing. Next, we compare these models across three datasets on the predictive performance on the purchase timing as well as purchase frequency. We show that while the Pareto/NBD and its Hierarchical Bayes extension [HB] models perform the best in predicting transaction frequency, the PDO and HB models predict transaction timing more accurately. Furthermore, we find that differences in a models predictive performance across datasets can be explained by the correlation between behavioral parameters and the proportion of customers without repeat purchases.

**Keywords**: Buy–till–you–defect models; purchase timing; Bayesian estimation; customer base analysis; probability models

**JEL codes**: C11; C41; M11

| **ERIM Report Series *Research in Management*** | |
|---|---|
| ERIM Report Series reference number | ERS–2013–001–LIS |
| Date of publication | 2013–01–08 |
| Version | 08–01–2013 |
| Number of pages | 49 |
| Persistent URL for paper | http://hdl.handle.net/1765/38235 |
| Email address corresponding author | ekorkmaz@rsm.nl |
| Address | Erasmus Research Institute of Management (ERIM)<br>RSM Erasmus University / Erasmus School of Economics<br>Erasmus University Rotterdam<br>PO Box 1738<br>3000 DR Rotterdam, The Netherlands<br>Phone: +31104081182<br>Fax: +31104089640<br>Email: info@erim.eur.nl<br>Internet: http://www.erim.eur.nl |
| Availability | The ERIM Report Series is distributed through the following platforms:<br>RePub, the EUR institutional repository<br>Social Science Research Network (SSRN)<br>Research Papers in Economics (RePEc) |
| Classifications | The electronic versions of the papers in the ERIM Report Series contain bibliographic metadata from the following classification systems:<br>Library of Congress Classification (LCC)<br>Journal of Economic Literature (JEL)<br>ACM Computing Classification System<br>Inspec Classification Scheme (ICS) |

# "Counting Your Customers": When will they buy next?

## An empirical validation of probabilistic customer base analysis models based on purchase timing

E. Korkmaz, R. Kuik, D. Fok

This research provides a new way to validate and compare buy-till-you-defect [BTYD] models. These models specify a customer's transaction and defection processes in a non-contractual setting. They are typically used to identify active customers in a company's customer base and to predict the number of purchases. Surprisingly, the literature shows that models with quite different assumptions tend to have a similar predictive performance.

We show that BTYD models can also be used to predict the timing of the next purchase. Such predictions are managerially relevant as they enable managers to choose appropriate promotion strategies to improve revenues. Moreover, the predictive performance on the purchase timing can be more informative on the relative quality of BTYD models.

For each of the established models, we discuss the prediction of the purchase timing. Next, we compare these models across three datasets on the predictive performance on the purchase timing as well as purchase frequency.

We show that while the Pareto/NBD and its Hierarchical Bayes extension [HB] models perform the best in predicting transaction frequency, the PDO and HB models predict transaction timing more accurately. Furthermore, we find that differences in a model's predictive performance across datasets can be explained by the correlation between behavioral parameters and the proportion of customers without repeat purchases.

**Keywords:** Buy-till-you-defect models, purchase timing, Bayesian estimation, customer base analysis, probability models.

# 1 Introduction

Many firms routinely store data on customer transactions. However, processing this data in order to provide managerially relevant information can still be a challenge. The customer base analysis literature provides a number of methods to use such data to gain a good understanding of the customer's transaction behavior. In the literature, a distinction is made between a contractual and a noncontractual setting. The latter is especially challenging as one does not observe the moment at which a customer leaves the company. In this setting, it is interesting to predict the number of future purchases, and to infer from observed behavior whether a customer has already left the company. A wide variety of models is available for these purposes.

The online retail industry is an important example of an industry operating in a noncontractual setting. Retailers never know which customers are *active*, or in other words, which customers will continue to buy from the firm. Thus, the customer database of an online retailer is likely to contain many inactive customers. For example, in October 2005, eBAY reported 168 million registered customers but only 68 million of them were counted as active by the company (Gupta et al. 2006). It is, therefore, very useful to develop a method to identify active customers under a noncontractual setting.

It has been widely recognized in the literature that models that ignore defection, like the early NBD model by Ehrenberg (1988), do not provide good predictions for this type of industry. They generally overestimate future transaction frequencies (Schmittlein and Peterson 1994). Schmittlein, Morrison, and Colombo (1987) proposed the first model that does account for defection. Since then, there has been a strong focus on the so-called buy-till-you-defect [BTYD] model. Several extensions of the model by Schmittlein, Morrison, and Colombo (1987) have been introduced (Fader, Hardie, and Lee 2005a, Abe 2009a and Jerath, Fader, and Hardie 2011). Some of these models have also been used to generate managerially relevant insights (Reinartz and Kumar 2000, Reinartz and Kumar 2003, and Wübben and Wangenheim 2008). However, little attention has been paid to providing a rigorous empirical comparison of the growing number of BTYD models. The models have mainly been compared on their performance in predicting a customer's number of purchases in a time interval.

In this paper, we suggest to include another measure in the comparison, namely the timing of the purchases. The existing models mainly differ in the distribution that governs the defection process. However, differences in the shape of this distribution may not directly lead to substantial differences in the expected number of purchases. Other measures, such as the customer being active at the end of the observation interval, directly involve the (unobserved) time of defection. If we want to use such

measures for validation, we require additional assumptions or heuristics. The timing of the purchase is, however, observed and critically depends on the interplay between its transaction and defection processes. Yet, predicting the timing of the next purchase is not straightforward. We develop methods for all state-of-the-art BTYD models. Based on these predictions, we provide an extensive empirical validation and comparison of these models where we go beyond the typical comparison that mainly considers purchase frequency.

We present the in-sample and out-of-sample performance on predicting the transaction frequency and the transaction timing of each customer for three datasets. The first dataset is from an online grocer in the Netherlands. The second is the well-known CDNOW dataset which has been commonly used as a benchmark set. The third dataset is also used by Batislam, Denizel, and Filiztekin (2007), and Jerath, Fader, and Hardie (2011) and is from a Turkish grocery retailer.

Our results show that different models can lead to different predictions on timing and frequency. It is important to understand how the underlying behavioral assumptions of the models lead to differences in performance. It turns out that certain data characteristics favor use of certain models.

The remainder of this paper is structured as follows. Section 2 gives an overview of the existing literature on BTYD models. We discuss the main features of and differences across the models, and present our contribution in more detail. In Section 3, we provide technical details of the considered models and present new results that deal with the timing of transactions. Section 4 gives a detailed description of the datasets. After presenting results of the empirical study in Section 5, general conclusions are discussed in Section 6.

## 2 BTYD Models

In this section, we briefly compare the main ideas underlying the BTYD models. We also discuss the similarities and differences across these models. Next, we review earlier empirical validation studies. Table 1 gives a summary of the related empirical work. We omit from this table studies that employ the Pareto/NBD model without testing its predictive performance in a holdout period (Reinartz and Kumar 2000, Reinartz and Kumar 2003 and Wu and Chen 2000). Finally, we discuss lifetime estimation using the models.

## 2.1 Models in comparison

The Pareto/Negative Binomial Distribution (Pareto/NBD) model (Schmittlein, Morrison, and Colombo 1987) is the first model that includes the customer's defection process. This model assumes that, while alive, customers make purchases according to a Poisson process with heterogeneous rates. The lifetime of a customer is modeled using an exponential distribution, also with a heterogeneous rate. The individual-specific rates of both processes are next treated as random effects and modeled using independent gamma distributions. This model allows for individual-level calculations on the probability of being active and the number of future purchases. The structure of the model leads to closed-form expressions for such predictions given the (hyper)parameters of the heterogeneity distributions. This feature has made this model useful for today's personalized marketing concepts such as direct marketing, one-to-one marketing and customer relation management.

Three important extensions of the Pareto/NBD model have been introduced in the literature. Fader, Hardie, and Lee (2005a) suggested replacing the continuous time defection process by a discrete time process. After each purchase, the customer defects with an individual-specific probability. The resulting model is called a Beta-Geometric/Negative Binomial Distribution (BG/NBD) model. The disadvantage of this model is that frequent purchasers have more "opportunities" to defect. In some cases this may not correspond to reality. To solve this problem, Jerath, Fader, and Hardie (2011) introduced the Periodic-Death-Opportunity (PDO) model. This model is very similar to the BG/NBD, but defection opportunities are defined in calendar time. In other words, defection can only occur at certain time intervals, independent of the transaction timing.

Another extension of the Pareto/NBD model deals with the relation between the purchase rate and the defection rate. In the Pareto/NBD model, and in the above-mentioned extensions, these rates are assumed to be independent. In practice, this assumption may be violated as, for example, frequent shoppers tend to have a long lifetime. This would imply a negative correlation between both rates. Abe (2009a) recently suggested a Hierarchical Bayes extension of the Pareto/NBD model that incorporates such correlation. In this model, the two gamma distributions are replaced by a bivariate log-normal distribution. Next to the possibility to capture correlations, another advantage of this model is that individual-specific covariates can be used. A disadvantage of this extension is that for some quantities, closed-form expressions are no longer available. As a result, the proposed model by Abe (2009a) needs Bayesian (simulation) techniques. We will refer to this model as the HB model.

## 2.2 Model performance

The first empirical validation study in the field, which reports the performance of a BTYD model in a holdout period, is presented by Schmittlein and Peterson (1994). This study not only provides an extensive empirical validation of the Pareto/NBD model, but also extends the model by adding the customer's spending decision. A major contribution of this paper is that it provides insights into the sampling properties of parameter estimates. For instance, the authors show how the accuracy of parameter estimation depends on the average observation time and on the number of customers in the sample (the space/time trade-off). Schmittlein and Peterson (1994) also examine whether customer characteristics can help in predicting transaction and defection behavior. In an application in the business-to-business market, they show that some groups of customers tend to have higher transaction rates while others have higher average dropout rates or a greater variation in dropout rates.

Fader, Hardie, and Lee (2005a) also include a validation study. This study compares the performance of the BG/NBD and the Pareto/NBD models on data from the online CD retailer CDNOW. They show that replacing the exponential dropout process (of Pareto/NBD) with a geometric one (BG/NBD) improves the model fit in the calibration period. The Pareto/NBD model, however, performs slightly better than the BG/NBD based on the quality of predictions of individual-level transactions in the forecast period. Fader, Hardie, and Lee (2005a) argue that the BG/NBD model is a good alternative for the Pareto/NBD model as it has similar performance, but requires fewer resources for parameter estimation.

In a third study, Batislam, Denizel, and Filiztekin (2007) compare the Pareto/NBD and BG/NBD models in terms of predicting the future number of transactions and the accuracy of the probability of being active. The comparison is based on loyalty card data from a specific store of a large grocery chain in Turkey. The authors also present a slight variation on the BG/NBD model. In this modified BG/NBD (MBG/NBD) model, customers may also drop out at time zero that is directly after making their first purchase. The MBG/NBD model yields almost identical estimates for the expected number of repeat purchases to the BG/NBD model. The general conclusion is that both the Pareto/NBD and the MBG/NBD models show similar performance on customer's purchase and defection processes.

Wübben and Wangenheim (2008) compare the Pareto/NBD and the BG/NBD models against managerial heuristics. In general, these heuristics are easy to implement, but are less detailed in terms of their predictions. Wübben and Wangenheim (2008) focus on predicting the number of future transactions and classifying active versus inactive customers. In terms of this classification, the man-

agerial heuristics perform at least as well as the models. However, the models perform better than the heuristics when predicting future transactions numbers. In this paper, the authors identify a potentially important problem of the BTYD models. On some datasets, the models produce extremely high probabilities of being active. Such high probabilities correspond to extremely long (residual) lifetime estimates.

Abe (2009a) compares his HB model to the Pareto/NBD model. He finds a similar fit and predictive performance. The disaggregate fit measures are the Mean Squared Error [MSE] of the predicted transaction numbers of individual customers, and the correlation between these predictions and the corresponding realizations. With regard to predicting future transaction numbers, the HB model performs slightly better than the Pareto/NBD model on two of the three datasets. The covariance matrix of the heterogeneity distribution is used to test the independence assumption of the Pareto/NBD. No significant dependency is found for any of the three datasets.

Finally, Jerath, Fader, and Hardie (2011) compare their PDO model to the Pareto/NBD and BG/NBD models using two datasets. They pay more attention to the defection process, and check model's performance on the median of lifetime estimates for each model. Note that the median lifetime is considered here, not the mean lifetime. Previous research has shown that the former is a better descriptor of the lifetime distribution (Reinartz and Kumar 2000) as using the median results in less extreme lifetime predictions. At a first glance, the Pareto/NBD and the PDO models produce similar results on the median lifetime. However, the PDO model predicts longer lifetimes for a randomly chosen customer than the Pareto/NBD model. The BG/NBD model's estimates are very different in that it predicts extremely long lifetimes. Based on these results, the authors suggest that the modeling of the defection process needs to be improved. Jerath, Fader, and Hardie (2011) also compare the models with respect to the predictions of the number of transactions. The Pareto/NBD and the PDO models show similar predictive performance and generally outperform the BG/NBD model.

## 2.3 Lifetime estimation

The BTYD models are usually compared on two dimensions: transaction frequency and lifetime related measures. Mostly, the first dimension is emphasized. An important challenge with the second dimension is that the exact lifetime is never observed. Even the state of a customer (active or inactive) can never be perfectly measured. There have been many attempts to validate predictions on customer lifetime or the active/inactive state. However, the majority of these studies acknowledge that the used

Table 1: Literature on the empirical validation and comparison of BTYD models

| Paper | Model(s) | Dataset(s) | Measures/Metrics | Results + Insights + Notes |
|---|---|---|---|---|
| Schmittlein, Peterson, (1994) | Pareto/NBD | 1 (a B2B office products supplier) | •Individual and aggregate level # future transaction<br>•Customer's active/inactive status<br>•Dollar volume of transactions | •Dollar volume of transactions is added to model.<br>•Customer's actual active status is designated by telephone interviews and significant evidence on model's ability to distinguish active customers is found.<br>•Sampling properties are added; # customers and observed time (T) tradeoff.<br>•Dropout process is validated by comparing to NBD model.<br>•Pareto/NBD performs better in predicting future transaction # than a simple heuristic. |
| Fader, Hardie, Lee, (2005) | Pareto/NBD - BG/NBD | 1 (CDNOW) | •Chi-square goodness-of-fit test<br>•Individual and aggregate level # future transaction | •The transition from exponential to geometric distribution improves model fit performance (without a significant loss in prediction power).<br>•BG/NBD is a good alternative to Pareto/NBD requiring fewer resources for parameter estimation. |
| Batislam, Denizel, Filiztekin, (2007) | Pareto/NBD - BG/NBD - MBG/NBD | 1 (a store of a large grocery retail chain) | •Chi-square goodness-of-fit test<br>•Individual and aggregate level # future transaction<br>•Active status of customers by computation of being active probability | •Pareto/NBD and MBG/NBD have similar estimates of # future transactions.<br>•Pareto/NBD model assigns slightly smaller active probabilities. |
| Wübben, Wangenheim, (2008) | Pareto/NBD - BG/NBD[a] | 3 (an apparel retailer, a global airline, CDNOW) | •Individual and aggregate level # future transaction<br>•Active status of customers | •Only Pareto/NBD model is compared against hiatus heuristic to distinguish active customers. Hiatus heuristic performs better. A sensitivity analysis on the threshold gives a similar result.<br>•Both BTYD models outperform heuristics in predicting # transactions |
| Abe, (2009) | Pareto/NBD - HB | 3 (CDNOW, a department store, music CD chain) | •Model fit assessment with correlation, MSE and MAPE<br>•Individual and aggregate level # future transaction | •Pareto/NBD and HB model show similar fit.<br>•The marginal log-likelihood suggests that HB model with covariates is better than without.<br>•HB model performs slightly better than Pareto/NBD model on 2 (out of 3) datasets.<br>•Independence assumption of Pareto/NBD has been examined and no significant dependency has been found. |
| Jerath, Fader, Hardie, (2011) | Pareto/NBD - BG/NBD - PDO | 2 (CDNOW, a store of a large grocery retail chain[b]) | •Chi-square goodness-of-fit test<br>•Individual and aggregate level # future transaction<br>•Median lifetime | •PDO shows significant improvement on calibration-period model fit compared to Pareto/NBD. Similar fit performance with BG/NBD model.<br>•Pareto/NBD and PDO models show similar predictive performance on # transactions and outperform BG/NBD model.<br>•Pareto/NBD and PDO show similar results on median lifetime whereas BG/NBD model predicts extreme lifetimes.<br>•PDO model suggests modeling of defection process can be improved. |

[a]Simple managerial heuristics are included in the comparison study.
[b]The same dataset from Batislam, Denizel, and Filiztekin (2007) is used.

indicators are not perfect.

Schmittlein and Peterson (1994) use telephone interviews to validate customer defection predictions. Customers are called and asked about their intentions to purchase from the company at an unspecified time in the future. However, even such a direct contact with a customer may not lead to the 'actual' defection information. It is known that customer's intentions are imperfect predictors of future behavior (Morwitz and Schmittlein 1992).

Batislam, Denizel, and Filiztekin (2007), Reinartz and Kumar (2000) and Wübben and Wangenheim (2008) base the 'true' active status of a customer on observed purchase activity in a holdout period. The model's predictive performance in terms of the defection process is next evaluated on this active status. However, as acknowledged by Wübben and Wangenheim (2008), customers who have not purchased in the holdout period may still be active and make a purchase after that period. In this sense, such a comparison is not fair and leads to favoring models that underestimate the lifetime. This is especially true, if the holdout period is short and/or the purchase rate is low.

Apart from the complexity of validating lifetime predictions, the managerial relevance of the lifetime concept has also been questioned. Reinartz and Kumar (2000) challenge the implicitly assumed strong association between lifetime and profitability in the noncontractual setting. Contrary to the general claim that a long customer lifetime is always desirable, they find that revenues mainly drive the lifetime value of a customer, not the duration of customer tenure. This argument is particularly valid in industries where customer switching costs are small (Reinartz and Kumar 2000). Furthermore, Jerath, Fader, and Hardie (2011) show that lifetime estimations from various BTYD models can vary to a large extent.

As aforementioned, in some cases, the BTYD models give extremely high active probabilities, which correspond to the extreme lifetime estimations (Wübben and Wangenheim 2008). Such clearly incorrect predictions could lead to a reluctance to use these models in practice. Perhaps with this in mind, Reinartz and Kumar (2000) strongly suggest firms not to neglect the transaction orientation of their business and to manage the short term accordingly.

## 2.4 Our contribution

Based on the discussion above, the only theoretically valid measure that is available to compare the BTYD models seems to be the accuracy of the predicted (future) transaction frequency. However, although the existing models are quite different in terms of their specification, they produce similar

predictions on this measure. In other words, this measure is not sensitive to differences among the models. In this paper, we introduce a new performance metric to overcome this problem.

Our measure is based on the timing of transactions and represents an observable value. Given the memoryless property on interarrival times of transactions in the considered BTYD models, we can predict the timing of the first and the last transaction in a certain period. As an in-sample metric, we propose the timing of the last in-sample transaction; as a holdout metric, we propose the minimum of the timing of the first out-of-sample transaction and the end of the holdout period.

In this paper, we compare the existing models on this new measure and on the predicted number of purchases. To make this possible, we derive formulas on the timing of transactions for each of the BTYD models. The methodology to calculate these timing predictions is also an important contribution of this paper. Besides providing a more rigorous comparison among BTYD models, these predictions also have managerial relevance. Predictions on the timing of the next purchase for each customer could be important information for both marketing and operations managers.

To our knowledge, our paper is the first to bring all the following models together: the Pareto/NBD, BG/NBD, the Hierarchical Bayes extension of the Pareto/NBD, and the recently proposed PDO model. Next, we are the first to compare these models based on the timing of purchases. A challenge in the comparison is that the models exhibit differences in their estimation procedures. The Pareto/NBD, BG/NBD and PDO models have closed-form expressions on some statistics for a 'randomly' chosen customer, such as the probability of being active and the expected number of future purchases. These models also yield closed form expressions for some statistics conditional on the observed transaction pattern of a customer. On the other hand, the HB model does not provide an analytical expression for important quantities due to the log-normal heterogeneity distribution. For this model, there is no closed-form expression for any relevant statistic not even for a randomly chosen customer. However, the complete distribution on any statistic can be obtained for each customer using MCMC methods. In order to overcome the difficulty of comparing the models, we bring the Pareto/NBD, BG/NBD and PDO models to the level of the HB model. More exactly, we obtain the complete individual-level distribution on the behavioral parameters for each model conditional on observed behavior. This provides great flexibility when computing various individual-level performance metrics.

# 3 Models and the Timing of Transactions

In this section, we present the models in technical terms. All models provide a representation of individual behavior by considering two arrival processes: one on purchase and one on defection. Individuals are assumed to make transactions according to the purchase process until they defect. The defection and transaction processes for individual $i$ depend on individual-specific parameters which we denote by $\theta_i$. On the population-level, all models specify a heterogeneity distribution for (the elements of) $\theta_i$. This distribution is parameterized by hyperparameters which are denoted by $\xi$. Below, we give the details for each model, and present expressions for the last transaction timing in the calibration period and the first transaction timing in the holdout period. The timing expressions vary depending on the assumptions of the models. To our knowledge, these expressions have not been presented before.

Table 2 gives a summary of the assumptions and the dominant estimation method for each model. We distinguish between assumptions on individual behavior and on heterogeneity. All models have the same assumption on the purchase process of an individual, while active. The models do differ in the defection process or in the heterogeneity distribution.

Table 2: Model comparison with respect to the assumptions and estimation process

| | Pareto/NBD | BG/NBD | PDO | HB |
|---|---|---|---|---|
| Purchase process | Poisson | Poisson | Poisson | Poisson |
| Defection process | Exponential | Shifted geometric | Shifted geometric | Exponential |
| Defection timing | Continuous | On purchase moments | Fixed periods | Continuous |
| Purchase rate distribution | Gamma | Gamma | Gamma | Bi-variate log-normal |
| Defection rate distribution | Gamma | Beta | Beta | |
| Estimated parameters | Hyperparameters | Hyperparameters | Hyperparameters | Hyper & individual par. |
| Estimation procedure | MLE | MLE | MLE | MCMC |

Before we present the models, we briefly discuss the general ideas used for calculating the predictions.

## 3.1 Conditional and unconditional inference

One can use the BTYD models to obtain predictions on different metrics. However, closed-form expressions for individual-level metrics conditional on the observed data are not always available. Below we indicate how to calculate such metrics. Suppose we want to predict a particular metric for customer $i$, we denote this as metric$_i$. There are two options: to include or not to include the purchase

history of this customer. The latter case is mainly relevant for in-sample predictions (model validation) and, the prediction can be seen as a prediction for a *randomly* chosen customer. We label this as unconditional inference. The former is relevant for out-of-sample predictions. These predictions are made conditional on data of the specific customer.

For conditional inference, we need to calculate $\mathbb{E}[\text{metric}_i|\text{all data}]$. We rewrite this expectation as

$$\begin{aligned}\mathbb{E}[\text{metric}_i|\text{all data}] &= \int_{\theta_i} \mathbb{E}[\text{metric}_i|\text{data}_i,\theta_i]\pi(\theta_i|\text{all data})\,\mathrm{d}\theta_i \\ &= \int_{\xi}\int_{\theta_i} \mathbb{E}[\text{metric}_i|\text{data}_i,\theta_i]\pi(\theta_i|\text{data}_i,\xi)\pi(\xi|\text{all data})\,\mathrm{d}\theta_i\,\mathrm{d}\xi,\end{aligned} \tag{1}$$

where $\theta_i$ denotes the individual-level parameters for individual $i$ and $\xi$ denotes the hyperparameters associated with the whole customer base in the focal BTYD model. In Sections 3.2 to 3.5, we provide closed-form expressions for $\mathbb{E}[\text{metric}_i|\text{data}_i,\theta_i]$ for each model. Calculating the integrals in (1) can still be very complex. However, samples from $\pi(\theta_i|\text{all data})$ can be obtained for all models. If the model relies on Maximum Likelihood Estimation [MLE], $\pi(\xi|\text{all data})$ is seen as a point mass at the Maximum Likelihood estimate $\hat{\xi}$, and draws are obtained by sampling from $\pi(\theta_i|\text{data}_i,\hat{\xi})$. For BG/NBD and PDO models, closed-form expressions are available for these conditional densities and we can apply direct sampling. For the other models, draws from the posterior are obtained using a Metropolis-Hastings MCMC sampler (Hastings 1970). In general, we approximate the integral for all models using

$$\mathbb{E}[\text{metric}_i|\text{all data}] \approx \frac{1}{L}\sum_{l=1}^{L}\mathbb{E}[\text{metric}_i|\text{data}_i,\theta_i^{(l)}],$$

where $\theta_i^{(l)}$, $l = 1,\ldots,L$, are draws from the posterior $\pi(\theta_i|\text{all data})$.

In the case of unconditional inference we need to calculate

$$\begin{aligned}\mathbb{E}[\text{metric}_i|\text{all data}_{-i}] &= \int_{\theta_i} \mathbb{E}[\text{metric}_i|\theta_i]\pi(\theta_i|\text{all data}_{-i})\,\mathrm{d}\theta_i \\ &= \int_{\xi}\int_{\theta_i} \mathbb{E}[\text{metric}_i|\theta_i]\pi(\theta_i|\xi)\pi(\xi|\text{all data}_{-i})\,\mathrm{d}\theta_i\,\mathrm{d}\xi \\ &\approx \int_{\xi}\int_{\theta_i} \mathbb{E}[\text{metric}_i|\theta_i]\pi(\theta_i|\xi)\pi(\xi|\text{all data})\,\mathrm{d}\theta_i\,\mathrm{d}\xi\end{aligned} \tag{2}$$

where $\text{all data}_{-i}$ denotes the available data ignoring the data for individual $i$. In the last line, we assume that enough data is available such that the contribution of a single individual to the conditional

distribution of the hyperparameters can be ignored. In this case we approximate the expectation by

$$\mathbb{E}[\text{metric}_i|\text{all data}_{-i}] \approx \frac{1}{L}\sum_{l=1}^{L}\mathbb{E}[\text{metric}_i|\theta_i^{(l)}].$$

If hyperparameters are estimated using MLE, $\theta_i^{(l)}$ denotes a draw from $\pi(\theta_i|\hat{\xi})$, with $\hat{\xi}$ the Maximum Likelihood estimate. If Bayesian estimation is used, the draws are obtained by first sampling $\xi^{(l)}$ from $\pi(\xi|\text{all data})$ and next sampling $\theta_i^{(l)}$ from $\pi(\theta_i|\xi^{(l)})$.

In the sections below, we present the expressions for the conditional expectation of the timing of the last in-sample transaction and the next out-of-sample transaction together with the sampling schemes for the behavioral parameters.

## 3.2 Pareto/NBD Model

In the Pareto/NBD model, customer $i$ remains active for a stochastic lifetime ($t_{\Delta,i}$) which has an exponential distribution with rate $\mu_i$. While active, this customer makes purchases according to a Poisson process with rate $\lambda_i$. The purchase rate and the defection rate are assumed to be distributed according to two independent gamma distributions across the population. The distribution for $\lambda_i$ has shape parameters $r$, and scale parameter $\alpha$. The shape and scale parameters for $\mu_i$ are $s$ and $\beta$, respectively.

The parameters of the heterogeneity distributions can be estimated by MLE. The likelihood can be written in terms of the number of purchases ($x_i$) and the timing of the last purchase ($t_{x,i}$) for each customer. This estimation procedure can be quite tedious from a computational perspective as the likelihood function involves numerous evaluations of the Gaussian hypergeometric function.

Schmittlein, Morrison, and Colombo (1987) presented some key expressions such as the probability of being active at the end of the calibration period ($T_i$) and the expected number of future transactions in a given time period for both a randomly chosen customer and a customer with past observed data ($x_i, t_{x,i}, T_i$).

The Pareto/NBD model allows us to predict also the timing of the last transaction in the calibration period and the timing of the first transaction in the holdout period. Given the individual-level parameters $\lambda_i$ and $\mu_i$, we can obtain the expected timing of the last purchase as

$$\mathbb{E}[t_{x,i}|\lambda_i,\mu_i,T_i] = \frac{1-e^{-\mu_i T_i}}{\mu_i} - \frac{1-e^{-(\lambda_i+\mu_i)T_i}}{\lambda_i+\mu_i}, \tag{3}$$

see Appendix A.1 for the associated derivations. By comparing $\mathbb{E}[t_{x,i}|\lambda_i,\mu_i,T_i]$, averaged over the estimated distribution of $\lambda_i$ and $\mu_i$, to the observed timing of the final purchase, we can assess the model's fit performance.

To measure the model's performance on out-of-sample predictions, we can use the timing of the first purchase in the interval $[T_i, T_i^+]$, where $T_i^+$ marks the end of the out-of-sample period. A complication here is that a particular customer may not make any purchase in this interval. For example, this may happen if the customer has defected. In turn, this makes it extremely difficult to compare the predictions to realizations. We solve this by instead predicting the minimum of the next purchase timing and $T_i^+$; for individual $i$ this minimum is denoted by $t_{f,i}$. If the customer has defected, $t_{f,i} = T_i^+$.

In Appendix A.1, we show that the conditional expectation of $t_{f,i}$ in the Pareto/NBD model equals

$$\begin{aligned}\mathbb{E}[t_{f,i}|x_i,t_{x,i},T_i,\lambda_i,\mu_i] = (1-\mathbb{P}[t_{\Delta,i}>T_i|x_i,t_{x,i},T_i,\lambda_i,\mu_i])T_i^+ \\ +\mathbb{P}[t_{\Delta,i}>T_i|x_i,t_{x,i},T_i,\lambda_i,\mu_i](T_i+\frac{1-e^{-(\lambda_i+\mu_i)(T_i^+-T_i)}}{\lambda_i+\mu_i}),\end{aligned} \quad (4)$$

where $\mathbb{P}[t_{\Delta,i}>T_i|x_i,t_{x,i},T_i,\lambda_i,\mu_i]$ gives the probability that individual $i$ is still active at time $T_i$. This probability can be shown to equal

$$\frac{\lambda_i}{\lambda_i+\mu_i e^{(\lambda_i+\mu_i)(T_i-t_{x,i})}}, \quad (5)$$

see Schmittlein, Morrison, and Colombo (1987). Note that this probability depends on the time between the last (in-sample) purchase and $T_i$. There is still a chance of defection in this period, but, given the data, a purchase is impossible in that interval.

### Sampling of the behavioral parameters for the Pareto/NBD Model

The joint posterior distribution of the behavioral parameters, $\theta_i = (\lambda_i,\mu_i)$, of the Pareto/NBD model is characterized by the likelihood function, the independent gamma priors on these parameters, and the (ML estimates of the) hyperparameters, $\xi = (\alpha,r,\beta,s)$:

$$\begin{aligned}\pi(\theta_i|\text{data}_i,\xi) &= \pi(\lambda_i,\mu_i|r,\alpha,s,\beta,x_i,t_{x,i},T_i) \\ &\propto f(x_i,t_{x,i},T_i|\lambda_i,\mu_i)g(\lambda_i|r,\alpha)h(\mu_i|s,\beta) \\ &\propto \frac{\lambda_i^{x_i}}{\lambda_i+\mu_i}(\mu_i e^{-(\lambda_i+\mu_i)t_{x,i}}+\lambda_i e^{-(\lambda_i+\mu_i)T_i})\frac{\alpha^r}{\Gamma(r)}\lambda^{(r-1)}e^{-\alpha\lambda}\frac{\beta^s}{\Gamma(s)}\mu_i^{(s-1)}e^{-\beta\mu_i}.\end{aligned} \quad (6)$$

As mentioned before, among the models that rely on MLE, the Pareto/NBD model is the only one that

does not have a standard distribution of individual parameters, $\pi(\theta_i|\text{data}_i, \xi)$. A Metropolis-Hastings algorithm can be used to sample from this posterior density. Details of this sampling algorithm are presented in Appendix B.

## 3.3 BG/NBD Model

The BG/NBD model replaces the continuous defection process of the Pareto/NBD model by a discrete process. Customers can now only drop out at the moment of a repeat transaction. This implies that the defection process is explicitly dependent on the purchase process.

Jerath, Fader, and Hardie (2011) argue that such a dependency may not be realistic, as heavy buyers eventually get more opportunities to drop out. However, the advantage of this model is that its parameters can be estimated more easily. The individual's purchase process is Poisson with intensity $\lambda_i \sim \Gamma(r, \alpha)$ like in the Pareto/NBD model. The dropout probability for individual $i$ is denoted by $p_i$ and follows a beta distribution with shape parameters $a$ and $b$. The hyperparameters of the BG/NBD model can be estimated using MLE.

Fader, Hardie, and Lee (2005a) present the expression for the expected number of (future) transactions of each customer, conditioned upon the hyperparameters. In Appendix A.2, we derive the expected timing of the last in-sample transaction and the next out-of-sample transaction. Again, we truncate the next future transaction timing to the end of the out-of-sample period ($T_i^+$). The expected timing of the last in-sample transaction equals

$$\mathbb{E}(t_{x,i}|T_i, \lambda_i, p_i) = \frac{1}{1-p_i}\left(\frac{1-e^{-\lambda_i p_i T_i}}{\lambda_i p_i} - \frac{1-e^{-\lambda_i T_i}}{\lambda_i}\right), \tag{7}$$

and the conditional expectation of the timing of the next transaction equals

$$\begin{aligned}\mathbb{E}(t_{f,i}|x_i, t_{x,i}, T_i, \lambda_i, p_i) = (1-\mathbb{P}[t_{\Delta,i} > T_i|x_i, t_{x,i}, T_i, \lambda_i, \mu_i])T_i^+ \\ + \mathbb{P}[t_{\Delta,i} > T_i|x_i, t_{x,i}, T_i, \lambda_i, \mu_i](T_i + \frac{1-e^{-\lambda_i(T_i^+ - T_i)}}{\lambda_i}).\end{aligned} \tag{8}$$

For this model, the conditional probability of being active at time $T_i$ equals

$$\mathbb{P}[t_{\Delta,i} > T_i|x_i, t_{x,i}, T_i, \lambda_i, \mu_i] = 1 - \delta_{t_{x,i}>0}\frac{p_i e^{\lambda_i(T_i - t_{x,i})}}{1-p_i+p_i e^{\lambda_i(T_i - t_{x,i})}},$$

where $\delta_{t_{x,i}>0}$ is a 0/1 indicator, which equals 1 if consumer $i$ made a repeat purchase.

### Sampling of the behavioral parameters for the BG/NBD Model

To sample the individual rate parameters of the BG/NBD model, we again make use of ideas from Bayesian statistics. Directly sampling from the joint conditional distribution of $\lambda_i$ and $p_i$ is not easy. However, we can derive the full conditional distributions of $\lambda_i$ and $p_i$. We, therefore, propose to use a Gibbs sampler which successively draws from the conditional distribution of $\lambda_i$ given $x_i, t_{x,i}, T_i$ and $p_i$, and the conditional distribution of $p_i$ given $x_i, t_{x,i}, T_i$ and $\lambda_i$. After convergence, this Markov Chain generates draws from the joint conditional distribution. Details of the derivations of both distributions are presented in Appendix B.2. The conditional density of the purchase rate $\lambda_i$ is

$$\pi(\lambda_i|x_i,t_{x,i},T_i,p_i)=\frac{\frac{p_i}{(t_{x,i}+\alpha)^{x_i+r}}}{\frac{p_i}{(t_{x,i}+\alpha)^{x_i+r}}+\frac{1-p_i}{(T_i+\alpha)^{x_i+r}}}\varphi_{x_i+r,t_{x,i}+\alpha}(\lambda_i)+\frac{\frac{1-p_i}{(T_i+\alpha)^{x_i+r}}}{\frac{p_i}{(t_{x,i}+\alpha)^{x_i+r}}+\frac{1-p_i}{(T_i+\alpha)^{x_i+r}}}\varphi_{x_i+r,T_i+\alpha}(\lambda_i), \quad (9)$$

where $\varphi_{x,\beta}$ is the density of a gamma distribution with shape parameter $x$ and rate parameter $\beta$. The conditional density of the defection probability $p_i$ equals

$$\pi(p_i|x_i,t_{x,i},T_i,\lambda_i)=\frac{a}{a+(b+x_i-1)e^{-\lambda_i(T_i-t_{x,i})}}\beta_{a+1,b+x_i-1}(p_i)+ \\ \frac{(b+x_i-1)e^{-\lambda_i(T_i-t_{x,i})}}{a+(b+x_i-1)e^{-\lambda(T_i-t_{x,i})}}\beta_{a,b+x_i}(p_i), \quad (10)$$

where $\beta_{a,b}$ is the density of a beta distribution with parameters $a$ and $b$. As the distributions are mixtures of gamma or beta distributions, respectively, sampling from these distributions is straightforward.

## 3.4 PDO Model

The most recent BTYD model is the Periodic Death Opportunity (PDO) model. This model is based on the BG/NBD model, but assumes that a customer may only defect after each $\tau$ periods of time. The defection process is, therefore, no longer linked to purchase occasions and heavy purchasers do not get more defection opportunities. Jerath, Fader, and Hardie (2011) show that the PDO model can be seen as a generalization of the Pareto/NBD and the NBD model. If $\tau$ becomes very small, the PDO model approaches the Pareto/NBD model. The PDO model collapses to the NBD model when $\tau$ exceeds the observation period, leaving no dropout possibility for customers.

More precisely, the PDO model assumes that the interpurchase time for individual $i$ has an exponential distribution with parameter $\lambda_i \sim \Gamma(r,\alpha)$. Customers may defect with a probability of $p_i$ after

each $\tau$ periods, where $p_i$ follows a beta distribution with parameters $a$ and $b$. The PDO model has four hyperparameters for the heterogeneity distributions and the additional *period length* parameter $\tau$. MLE can again be used to estimate the hyperparameters; for more details see Jerath, Fader, and Hardie (2011).

The introduction of the $\tau$ parameter complicates the prediction of the timing of the last and the next transactions. $T_i$ is likely not a multiple of $\tau$, and we need to deal with the delay between the last opportunity to defect before $T_i$ and, for the computation of the expected first future transaction, the delay between $T_i$ and the first opportunity to defect after $T_i$. A further complication is the possibility that there is no defection opportunity during $(T_i, T_i^+]$. Details of the derivations are presented in Appendix A.3. The expected time of the last transaction in the in-sample period is

$$\mathbb{E}(t_{x,i}|T_i,\lambda_i,p_i) \;=\; \sum_{n=1}^{N_i} p_i(1-p_i)^{n-1}\left(n\tau - \frac{1-e^{-n\lambda_i\tau}}{\lambda_i}\right) + (1-p_i)^{N_i}\left(T_i - \frac{1-e^{-\lambda_i T_i}}{\lambda_i}\right), \quad (11)$$

where $N_i$ equals the number of defection opportunities, that is, $N_i = \lfloor T_i/\tau \rfloor$. The expected time of the first purchase in the out-of-sample period $(T_i, T_i^+]$ is

$$\begin{aligned}\mathbb{E}(t_{f,i}|x_i,t_{x,i},T_i,\lambda_i,p_i,T_i^+) = (1-p_i^+)T_i^+ + p_i^+\Bigg[&(T_i+\frac{1}{\lambda_i})e^{-\lambda_i T_i} - (\bar{T}_i+\frac{1}{\lambda_i})e^{-\lambda_i\bar{T}_i} + \delta_{T_i^+<(N_i+1)\tau}\,T_i^+ e^{-\lambda_i(T_i^+-T_i)} \\ &+\delta_{T_i^+\geq(N_i+1)\tau}\left(e^{-\lambda_i((N_i+1)\tau-T_i)}p_i T_i^+ + (1-p_i)\Big((N_i+1)\tau + \mathbb{E}(t^+|\lambda_i,p_i,T_i^+-(N_i+1)\tau)\Big)\right)\Bigg], \quad (12)\end{aligned}$$

where $\bar{T}_i$ is the minimum of the first defection opportunity in the out-of-sample period for customer $i$ and $T_i^+$, that is, $\bar{T}_i = \min((N_i+1)\tau, T_i^+)$. Furthermore, $p_i^+$ is shorthand notation for the conditional probability that individual $i$ is active at time $T_i$. This probability is given by

$$p_i^+ = \mathbb{P}(t_{\Delta,i} > T_i|x_i,t_{x,i},T_i,\lambda_i,p_i) = \frac{(1-p_i)^{N_i}e^{-\lambda_i T_i}}{p_i e^{-\lambda_i\tau}\sum_{n=m_{x,i}}^{N_i}\left((1-p_i)e^{-\lambda_i\tau}\right)^{n-1} + (1-p_i)^{N_i}e^{-\lambda_i T_i}}$$

where $m_{x,i}$ is the first opportunity to defect *after* $t_{x,i}$, that is, $m_{x,i} = \left\lfloor \frac{t_{x,i}}{\tau} + 1 \right\rfloor$ and we define $\sum_{n=a}^{b}(\cdot) = 0$ whenever $a > b$. Finally, $\mathbb{E}(t^+|\lambda_i,p_i,T_i^+-(N_i+1)\tau)$ is the expected value of the minimum of the time of the first transaction in $(0, T_i^+-(N_i+1)\tau)$ and $(T_i^+-(N_i+1)\tau)$. The expression for this expectation is given in Equation (36) of the appendix.

### Sampling of the behavioral parameters for the PDO Model

To sample $\lambda_i$ and $p_i$, we again propose a Gibbs sampler; see Appendix B.3 for the details. Conditional on the data and $p_i$, $\lambda_i$ follows a mixture of gamma distributions, that is,

$$\pi(\lambda_i|x_i,t_{x,i},T_i,p_i)=\sum_{n=m_{x,i}}^{N_i}\frac{w_{x_i,p_i}^{(n)}}{W_{x_i,t_{x,i},p_i}}\varphi_{x_i+r,\alpha+(n-1)\tau}(\lambda_i)+\frac{w_{x_i,p_i}^{(N_i+1)}}{W_{x_i,t_{x,i},p_i}}\varphi_{x_i+r,\alpha+T_i}(\lambda_i) \tag{13}$$

where $W_{x_i,t_{x,i},p_i}=\sum_{n=m_{x,i}}^{N_i+1}w_{x_i,p_i}^{(n)}$, and

$$w_{x_i,p_i}^{(n)}=\begin{cases}p_i\frac{(1-p_i)^{n-1}}{(\alpha+(n-1)\tau)^{x_i+r}} & \text{if } 1\le n\le N_i\\ \frac{(1-p_i)^{N_i}}{(\alpha+T_i)^{x_i+r}} & \text{if } n=N_i+1\,.\end{cases}$$

The conditional distribution of $p_i$ is a mixture of beta distributions, that is,

$$\pi(p_i|x_i,t_{x,i},T_i,\lambda_i)=\sum_{n=m_{x,i}}^{N_i}\frac{v_\lambda^{(n)}}{V_{t_{x,i},\lambda_i}}\beta_{a+1,b+n-1}(p_i)+\frac{v_{\lambda_i}^{(N_i+1)}}{V_{t_{x,i},\lambda_i}}\beta_{a,b+N_i}(p_i) \tag{14}$$

where $V_{t_{x,i},\lambda_i}=\sum_{n=m_{x,i}}^{N_i+1}v_{\lambda_i}^{(n)}$, and

$$v_{\lambda_i}^{(n)}=\begin{cases}B(a+1,b+n-1)e^{-\lambda(T_i-(n-1)\tau)} & \text{if } m_{x,i}\le n\le N_i\\ B(a,b+N_i) & \text{if } n=N_i+1,\end{cases}$$

where $B(\cdot,\cdot)$ is the beta function. Note that the value $V_{t_{x,i},\lambda_i}$ depends on the data only through $m_{x,i}$.

## 3.5 Hierarchical Bayes Extension of the Pareto/NBD Model

The models presented above do not allow the individual-level parameters to be correlated and they do not take into account customer characteristics. In many cases, individual-level characteristics are available and may be useful in predicting customer behavior. Abe (2009a), therefore, proposes a Hierarchical Bayes [HB] extension of the Pareto/NBD model in which the individual-level parameters follow a bivariate log-normal distribution. The mean of this distribution may depend on customer characteristics.

The disadvantage of this extension is that closed-form expressions for interesting metrics, such as the expected number of purchases, are no longer available. Besides, MLE can no longer be straight-

forwardly used to obtain parameter estimates. Abe proposes the use of Markov chain Monte Carlo [MCMC] techniques to estimate the (hyper)parameters and to calculate various metrics.

Abe (2009a) makes the same individual-level assumptions as in the Pareto/NBD model, but assumes that $(\log\lambda_i, \log\mu_i) \sim N(w_i\beta, \Gamma)$, where $w_i$ is a $1 \times K$ vector of individual characteristics, including an intercept. In case no covariates are available, the distribution reduces to $N(\beta, \Gamma)$. $\Gamma$ is not restricted to a diagonal matrix and, therefore, this model allows the individual-level parameters to be correlated.

The joint density of the data and all parameters forms the basis for the inference. This density is given by

$$\pi(\{x_i, t_{x,i}, T_i, \lambda_i, \mu_i\}_{i=1}^N, \beta, \Gamma) = \prod_{i=1}^{N} \Big( \pi(x_i, t_{x,i}|\lambda_i, \mu_i)\pi(\lambda_i, \mu_i|\beta, \Gamma) \Big) \pi(\beta, \Gamma)\,.$$

Here $\pi(\beta, \Gamma)$ is the prior distribution of the population-level parameters $\beta$ and $\Gamma$. The standard conjugate prior is used, that is, $\beta \sim N(\beta_0, A_o)$ and $\Gamma$ follows an inverted Wishart distribution with parameters $(\nu_0, \Gamma_0)$. As the individual-level behavioral assumptions of the HB model are identical to the Pareto/NBD model, conditional on $\lambda_i$ and $\mu_i$, all timing related expressions are the same. Draws for the individual-level parameters are a natural by-product of the MCMC sampler.

Abe (2009b) proposes an extension of the HB model by adding the amount of spending. Hereby, the individual parameter vector, $\theta_i$, extends to three dimensions, including the rate of average log-spending of customers, $(\log\lambda_i, \log\mu_i, \log\eta_i)$. We also include this extension in our empirical study. Consequently, we consider four different configurations of the HB model. The first configuration (HB1) represents the HB model without any covariates and without spending. The second configuration (HB2) incorporates only the customer-specific covariates. The third and fourth configurations represent the HB models with the average spending parameter, and without or with covariates, respectively.

### Sampling of the hyperparameters and the behavioral parameters for the HB Model

We use MCMC for inference on the hyperparameters and the individual parameters for the HB models. More specifically, we use a Metropolis within Gibbs sampler (see Hastings (1970)). The sampler uses the latent variables $z_i$ and $t_{\delta,i}$, where $z_i$ is the binary variable representing whether customer $i$ is active ($z_i = 1$) or inactive ($z_i = 0$) at the end of the calibration period; and if already inactive, $t_{\delta,i}$ is the defection time (see Abe (2009a)). As our sampler differs from the one presented in Abe (2009a), we present the main steps of the sampler:

[0] Set initial value for $\theta_i, i = 1, \ldots, N$.

[1a] Generate $z_i|t_{x,i}, x_i, T_i, \theta_i$ according to the being active probability given in Equation (5), for $i = 1, \ldots, N$.

[1b] If $z_i = 0$, generate $t_{\delta,i}|t_{x,i}, x_i, T_i, z_i, \theta_i$ using an exponential distribution truncated to $(t_{x,i}, T_i)$.

[2] Generate $\beta, \Gamma|\{\theta_i\}_{i=1}^N$ using a standard multi-variate normal regression update (see Rossi, Allenby, and McCulloch (2005, Page 34)).

[3] Generate $\theta_i|t_{x,i}, x_i, T_i, z_i, t_{\Delta,i}, \beta, \Gamma$ with a Gaussian random-walk MH algorithm, for $i = 1, \ldots, N$.

The step size in the random-walk MH algorithm is set by applying an adaptive MH method in the burn-in phase (Gilks, Richardson, and Spiegelhalter 1996).

# 4 Data

We compare the performance of the presented models on three datasets. Below, we briefly discuss these three datasets.

The first dataset contains daily transaction data of an online grocery retailer in the Netherlands. We base our analysis on a random set of 1460 customers who started buying from the company in January 2009. We ignore all Sundays as the company does not provide delivery on that day. The available data contains the initial and the repeat purchase information of each customer over a period of 309 days. To estimate the model parameters, we use the transaction data of all customers over the first 154 days, leaving a 155 day holdout period for model validation.

The second dataset is the commonly used CDNOW data. This publicly available dataset covers the transactions data of 2357 customers who made their first transaction in the first quarter of 1997. The data spans a period of 78 weeks from January 1997 through June 1998. We set the calibration and holdout periods to 39 weeks each.

The final dataset comes from a Turkish grocery store. This set is also used by Batislam, Denizel, and Filiztekin (2007) and Jerath, Fader, and Hardie (2011). It contains the transactions of 5479 customers who made their first purchase between August 2011 and October 2011, covering a period of 91 weeks. To be consistent with the earlier papers, we use the first 78 weeks for calibration and leave 13 weeks for validation purposes. Detailed descriptive statistics of all datasets appear in Table 3.

The three datasets have quite different characteristics. Together they span a wide range of purchase and activity patterns. For instance, in the first dataset, the majority of customers are frequent

Table 3: Descriptive statistics over the three datasets

| | Online grocer | CDNOW | Grocer |
|---|---|---|---|
| Number of customers | 1460 | 2357 | 5479 |
| Available time frame | 309 days | 78 weeks | 91 weeks |
| Time split (in-sample/out-of-sample) | 154/155 | 39/39 weeks | 78/13 weeks |
| Available time units | days | weeks/days | weeks |
| Zero repeaters in estimation period (fraction) | 174 (0.12) | 1,411 (0.60) | 2,221 (0.41) |
| Zero repeaters in holdout period (fraction) | 295 (0.20) | 1,673 (0.70) | 4,577 (0.84) |
| Zero repeaters in estimation and holdout periods (fraction) | 135 (0.09) | 1,218 (0.51) | 2,179 (0.40) |
| Number of purchases in estimation period (all) | 16,252 | 2,457 | 24,840 |
| Number of purchases in holdout period | 12,827 | 1,882 | 2,907 |
| Average number of purchases | | | |
| per customer in estimation period (stdev) | 11.13 (10.76) | 1.04 (2.190) | 4.53 (9.17) |
| Average number of purchases | | | |
| per customer in holdout period (stdev) | 8.79 (10.78) | 0.798 (2.057) | 0.53 (1.72) |
| Average length of the observation period ($T$) (stdev) | 143.76 (7.39) | 32.72 (3.33) | 22.81 (26.87) |
| Average recency as a fraction of $T$ ($(T-t_x)/T$) | 0.27 | 0.79 | 0.67 |

customers, whereas the other two datasets include a large group of incidental buyers. Although the first two datasets both deal with online retailers, the industries in which these retailers operate are different, namely groceries versus CDs. We see a clear difference in the customer's loyalty to the firm; the average frequency of shopping per customer is higher at the grocery retailer than at the CD retailer. The fraction of customers without a repeat purchase (zero-repeat buyers) is also much smaller for the online grocer compared to CDNOW. A customer's final observed purchase tends to be close to the end of the sample for the online retailer. This is reflected in the last row of Table 3, which gives the average recency normalized by the average observation period.

Customer behavior at the brick-and-mortar grocer is quite different compared to that at the online grocer. Contrary to the general claim in the literature, the customers of the online grocer are more loyal to the company than those of the grocer chain. The rate of zero-repeat buyers in the grocer's data base is considerably higher, and the average normalized recency is significantly lower than for the online grocer. In what follows, we relate the performance of the models on three datasets to their characteristics.

# 5 Empirical Findings

We split this section in two parts. First, we discuss the parameter estimates for all models and datasets[1]. Next, we focus on the predictive performance of the models, where we distinguish between (1) expected number of transactions; and (2) expected timing of transactions. We especially focus on the performance of the models in predicting the timing of the last in-sample purchase and the first out-of-sample purchase.

For the online retailer datasets (online grocer and CDNOW), covariate data on the average number of shopping items per customer is available. This data is used in the HB model configurations HB2 and HB4. As both datasets also have individual-level spending information, the spending extension of the HB models (HB3 and HB4) can be applied as well. We mean-center the covariate (average number of items in the shopping basket) so that the mean of the behavioral parameters, $\theta_i$, given average covariate values will be entirely determined by the intercept. As no covariate nor spending information is available for the third dataset (grocer), only the HB1 model can be used. For all HB models, the MCMC steps were repeated 256,000 iterations, of which the last 32,000 were used to infer the posterior distribution of parameters. Convergence was monitored visually and checked with the Geweke test on all datasets (Geweke et al. 1991).

## 5.1 Parameter estimates

### Maximum Likelihood-based models

First we present the parameter estimates that are based on ML estimation; namely for the Pareto/NBD, BG/NBD, and PDO models. Using the estimates, we can get insight in the degree of heterogeneity in each customer base as well as in some key quantities for a random customer. Table 4 reports the estimated hyperparameters for the online grocer. According to the Pareto/NBD model a random customer makes 0.072 transactions per day while active. Note that this statistic cannot be calculated directly from the data as it intrinsically contains the condition of being active. The shape parameter ($r = 0.958$) indicates a moderate level of heterogeneity in purchase rates across customers (Schmittlein, Cooper, and Morrison 1993). For this dataset, the PDO model fits best when the period length $\tau$ is set to about 20 days. The parameters related to the purchase process in the PDO model are very similar to those in the Pareto/NBD model. The BG/NBD model also gives a very similar result for the purchase

[1] All calculations are performed using MATLAB R2011b.

rate of an average customer while active (0.071 purchases). The relatively small shape parameter value ($r = 0.897$) indicates slightly more differences in purchase rates across customers within the BG/NBD model.

Table 4: Results of the Pareto/NBD, BG/NBD and PDO Maximum Likelihood Estimates - online grocery retailer

| Pareto/NBD | | BG/NBD | | PDO ($\tau = 20.001$) | |
|---|---|---|---|---|---|
| r | 0.96 | r | 0.90 | r | 0.94 |
| $\alpha$ | 13.35 | $\alpha$ | 12.64 | $\alpha$ | 13.13 |
| r/$\alpha$ | 0.072 | r/$\alpha$ | 0.071 | r/$\alpha$ | 0.071 |
| s | 0.04 | a | 0.03 | a | 0.04 |
| $\beta$ | 38.24 | b | 3.00 | b | 2.18 |
| s/$\beta$ | 0.001 | a/(a+b) | 0.010 | a/(a+b) | 0.018 |
| log-likelihood | -49,208 | log-likelihood | -49,212.3 | log-likelihood | -49,201.4 |

The estimated average defection rate for the Pareto/NBD model is given by $s/\beta = 0.001$. As the shape parameter $s$ is less than 1, the expected lifetime value of a random customer from the cohort diverges to infinity. From another perspective, half of the customers in the cohort defect after $(2^{1/s} - 1)\beta = 383,014,675$ days. This shows that a short-term measure rather than these long lifetime estimations would be more useful for a manager. The probability of a random customer defecting in the next day is only $1 - e^{-s/\beta} = 0.001$. In other words, it is highly unlikely that such a customer will drop out in the near future. However, the very small value of $s$ suggests that there is a very large dispersion in defection rates.

The estimation results for the CDNOW data are given in Table 5. We obtain the same parameter estimates as Fader, Hardie, and Lee (2005a). We find that an average customer makes around 0.05 transactions per week, while active. The small shape parameter value indicates substantial differences in purchase rates across customers. Similar to the previous dataset, the heterogeneity on defection rates is extremely high on this dataset ($s = 0.606$ in the Pareto/NBD model) and the expected lifetime value of a random customer from the cohort diverges to infinity.

When applying the models on the Turkish grocery dataset, we find that while active, an average customer places approximately 0.1 orders per week; see Table 6. The population is quite heterogeneous in purchase rates. The heterogeneity is even greater according to the BG/NBD model. For an in-depth discussion on the customer lifetime, we recommend the discussion in Jerath, Fader, and Hardie (2011).

Table 5: Results of the Pareto/NBD, BG/NBD and PDO Maximum Likelihood Estimates - CDNOW

| Pareto/NBD | | BG/NBD | | PDO ($\tau = 3.001$) | |
|---|---|---|---|---|---|
| r | 0.55 | r | 0.24 | r | 0.52 |
| $\alpha$ | 10.58 | $\alpha$ | 4.41 | $\alpha$ | 10.40 |
| r/$\alpha$ | 0.052 | r/$\alpha$ | 0.055 | r/$\alpha$ | 0.05 |
| s | 0.61 | a | 0.79 | a | 0.43 |
| $\beta$ | 11.66 | b | 2.43 | b | 2.61 |
| s/$\beta$ | 0.052 | a/(a+b) | 0.246 | a/(a+b) | 0.142 |
| log-likelihood | -9,595 | log-likelihood | -9,582.4 | log-likelihood | -9,585.6 |

Table 6: Results of the Pareto/NBD, BG/NBD and PDO Maximum Likelihood Estimates - grocery retailer

| Pareto/NBD | | BG/NBD | | PDO ($\tau = 1.001$) | |
|---|---|---|---|---|---|
| r | 0.48 | r | 0.28 | r | 0.46 |
| $\alpha$ | 4.38 | $\alpha$ | 2.34 | $\alpha$ | 4.38 |
| r/$\alpha$ | 0.11 | r/$\alpha$ | 0.12 | r/$\alpha$ | 0.105 |
| s | 0.57 | a | 0.40 | a | 0.62 |
| $\beta$ | 17.60 | b | 2.09 | b | 22.19 |
| s/$\beta$ | 0.033 | a/(a+b) | 0.161 | a/(a+b) | 0.027 |
| log-likelihood | -67,925.8 | log-likelihood | -68,008.3 | log-likelihood | -67,757.3 |

### MCMC-based models

In order to apply the HB models we first need to set the prior distributions. In many contexts, the prior is set diffuse enough so that it does not affect the posterior. In other words, the prior variance is set to a very large value. For the prior on $\Gamma$, we initially use $\nu_0 = J + 3$ and $\Gamma_0 = \nu_0 I$, where $J$ represents the number of behavioral parameters of a customer (see Rossi, Allenby, and McCulloch (2005, Page 30)). This is an extremely spread prior. However, in case limited data per individual is available, such a prior may have a strong impact on the posterior. Indeed, looking at the likelihood function for the HB model given in Equation (37), it can be seen that the likelihood for a zero-repeat buyer ($x_i = 0 = t_{x,i}$) tends to 1 as $\mu$ approaches $\infty$ for any value of $\lambda$. Therefore, without a proper prior the posterior does not exist. The prior needs to ensure that the posterior density for large values for $\mu$ approaches 0 quickly enough. Very diffuse priors fail to deliver this property, leading to (very) unstable estimates.

Among the datasets in our study, the CDNOW dataset is unique in terms of having a very large proportion of zero-repeat buyers. In other words, the data does not provide much information. We, therefore, need to set a relatively informative prior for this dataset. Accordingly, we choose $\nu_0 = J + 30$ and $\Gamma_0 = \nu_0 I$. In this way, extreme estimates are avoided and population-level estimates

are reasonable[2]. Still, we have experimented with a diffuse prior on this dataset. A detailed look at the results per individual (not reported) reveals that there are indeed extreme values for some parameters (in a range of $5.10^8$). We also observe very different predictions for individuals with a history of zero-repeat transactions, following the reasoning stated above. A further elaboration on the selection of the prior parameters on the CDNOW dataset is given in Appendix C.

The hyperparameters of the HB models are not directly comparable to the hyperparameters of the other BTYD models, not only because of the different heterogeneity distribution (log-normal distribution versus gamma and beta distributions), but also because the multi-variate structure of the log-normal distribution allows correlation between parameters for a single customer. Table 7 gives the median and the mode of the posterior mean of behavioral parameters across customers in each dataset. It is interesting to note that the location of the population distribution in the HB models seems to be different to that for the other models. In the next section, we investigate whether this has an impact on the models' performance.

Table 7: Median and mode of the behavioral rates of HB model estimates

| | | HB1 | | HB2 | | HB3 | | HB4 | |
|---|---|---|---|---|---|---|---|---|---|
| | | $\lambda$ | $\mu$ | $\lambda$ | $\mu$ | $\lambda$ | $\mu$ | $\lambda$ | $\mu$ |
| online grocer | median | 0.0474 | 0.0008 | 0.0471 | 0.0008 | 0.0479 | 0.0002 | 0.0479 | 0.0003 |
| | mode | 0.0204 | 0.0003 | 0.0233 | 0.0004 | 0.0086 | 0.0001 | 0.0085 | 0.0001 |
| CDNOW | median | 0.0045 | 0.0129 | 0.0072 | 0.0170 | 0.0081 | 0.3834 | 0.0089 | 0.5117 |
| | mode | 0.0045 | 0.0132 | 0.0073 | 0.0019 | 0.0080 | 0.0006 | 0.0083 | 0.0004 |
| grocer | median | 0.0469 | 0.0568 | - | - | - | - | - | - |
| | mode | 0.0464 | 0.0080 | - | - | - | - | - | - |

## 5.2 Unconditional predictions

We follow the procedure described in Section 3.1 to obtain unconditional predictions. As individuals in the customer database make their first purchases at different times, the time span $T$ varies across customers. Consequently, we obtain different in-sample predictions for different values of $T$. We calculate the unconditional predictions for each of the $T_i$ values in the database and average over them. These predictions are only based on the population-level parameters, estimated using all the data in the customer base. Hence, they serve as good indicators of the model's ability to fit the overall data pattern. Table 8 shows some statistics on the unconditional expectations on the number of

[2]With a more diffuse prior, an extremely large number of iterations is needed to obtain accurate estimates of posterior quantities as the posterior variance will be very large.

transactions and the timing of the last transaction for each model and each dataset. The first row shows the statistics based on the observed values for each dataset.

The mean predictions for the HB models are very different from the other model predictions on CDNOW data[3]. However, the predicted values are much closer to the median and mode of the data. In other words, it seems that the large number of zero-repeat buyers pulls the predictions from the HB models towards smaller values. This is probably due to the shape of the population distribution. As can be seen in Table 4, the mode for the population distributions of $\lambda_i$ and $\mu_i$ are at 0. The log-normal distribution does not allow for a mode at 0 without also pulling the mean towards 0 (or having an extreme variance). This explains why the mean predictions for the HB models are pulled towards 0. For the other datasets, the percentage of zero-repeat buyers is not as large, therefore this phenomenon is not observed there.

Table 8: Average of unconditional expectations versus observed quantities in calibration period

| | | Number of transactions | | | Time of last transaction | | |
|---|---|---|---|---|---|---|---|
| | | mean | median | mode | mean | median | mode |
| online grocer | True | 10.132 | 6 | 0 | 105.421 | 128 | 0 |
| | Pareto/NBD | 7.926 | 8.000 | 8.300 | 76.786 | 77.831 | 78.410 |
| | BG/NBD | 6.593 | 6.647 | 6.970 | 57.841 | 58.571 | 61.670 |
| | PDO | 9.789 | 9.884 | 10.360 | 104.217 | 105.574 | 111.540 |
| | HB1 | 10.573 | 10.694 | 11.150 | 103.157 | 104.419 | 110.650 |
| | HB2 | 10.707 | 10.826 | 11.320 | 106.048 | 107.289 | 113.780 |
| | HB3 | 11.231 | 11.341 | 11.290 | 101.139 | 102.482 | 107.830 |
| | HB4 | 11.139 | 11.256 | 11.360 | 101.662 | 102.942 | 104.270 |
| CDNOW | True | 1.042 | 0 | 0 | 6.864 | 0 | 0 |
| | Pareto/NBD | 1.071 | 1.071 | 1.100 | 6.804 | 6.790 | 6.860 |
| | BG/NBD | 1.058 | 1.057 | 1.000 | 6.913 | 6.889 | 7.760 |
| | PDO | 1.079 | 1.078 | 1.150 | 6.915 | 6.900 | 6.540 |
| | HB1 | 0.227 | 0.227 | 0.220 | 2.884 | 2.862 | 3.090 |
| | HB2 | 0.245 | 0.244 | 0.230 | 3.020 | 2.997 | 2.590 |
| | HB3 | 0.232 | 0.231 | 0.220 | 2.900 | 2.880 | 3.410 |
| | HB4 | 0.235 | 0.235 | 0.220 | 2.953 | 2.926 | 2.690 |
| grocer | True | 4.534 | 1 | 0 | 22.805 | 7 | 0 |
| | Pareto/NBD | 4.462 | 4.443 | 4.320 | 22.589 | 22.411 | 21.850 |
| | BG/NBD | 4.240 | 4.222 | 4.150 | 23.951 | 23.731 | 23.000 |
| | PDO | 4.424 | 4.403 | 4.290 | 22.841 | 22.667 | 22.110 |
| | HB1 | 4.839 | 4.816 | 4.700 | 22.485 | 22.313 | 21.910 |

We also provide some performance measures for the number of in-sample transactions ($x$) and the time of the last in-sample transaction ($t_x$) for each model. Table 9 shows the in-sample Mean Squared Error (MSE), Mean Absolute Error (MAE) on all predictions and Mean Error on the over- (ME+) and underpredicted (ME−) observations for all models on the three datasets. At a first glance, all models

[3]Note that the mean unconditional predictions move even further away with the most diffuse prior. For example, it becomes 0.09 for the HB2 model, see Table 15.

have a similar fit when predicting $x$. The PDO model performs slightly better with respect to MSE on the CDNOW and the grocery data. The estimated hyperparameters for this model lead to a low probability of extreme values on these datasets. On the other hand, the HB model fits the best in terms of MSE on the online grocery dataset. In terms of absolute errors in the unconditional predictions of $x$, the BG/NBD model has the best fit for the online grocer and the grocer data.

The HB models perform well on the CDNOW dataset in terms of the MAE. The high MSE and the low MAE values for the HB models on CDNOW link back to our earlier discussion. The high number of zero-repeat buyers in this dataset causes the predictions to move towards the mode of the data. Consequently, on this dataset, the mean of the unconditional predictions of the HB models approaches the strong mode of the data. This fact leads to a low MAE for the HB models. All models show an asymmetry in the unconditional prediction error. If the forecast is too high, the error tends to be relatively small.

The Pareto/NBD, BG/NBD and PDO models have a very similar performance when predicting the last purchase time on the CDNOW dataset. The PDO and the HB are the best performing models with respect to the unconditional predictions on this measure for the CDNOW and the online grocer datasets (considering the MSE and the MAE, respectively). On the grocer dataset, all models have a similar fit on predicting $t_x$, except the BG/NBD model which fits slightly worse on this metric.

Among the different configurations of HB models, we see that inclusion of covariates generally causes a slight increase in model fit on both measures. On the other hand, adding the spending parameter into the estimation procedure leads to a slight decrease in model fit for the frequency and the timing of in-sample transactions on the online grocer data.

## 5.3 Conditional predictions

In this section, we consider individual-level predictions conditional on the individual's history. As discussed in Section 3.1, for some metrics of interest, obtaining closed-form expression conditioned on an individual's history and hyperparameters can be extremely cumbersome because of the integral in Equation (2). We, therefore, first obtain draws for the individual's behavioral parameters from the posterior densities and next calculate the expected value of the metrics of interest by averaging over these draws. For the Pareto/NBD model, we use a Gaussian random-walk MH sampler to obtain draws of individual parameters conditional on the hyperparameters. To satisfy convergence, we repeat the

Table 9: In-sample predictive performance for unconditional predictions of the number of transactions ($x$) and the time of last transaction ($t_x$)

| | | $x$ | | | | $t_x$ in weeks | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | MSE | MAE | ME+ | ME− | MSE | MAE | ME+ | ME− |
| online grocer | Pareto/NBD | 116.636 | 7.803 | 4.847 | 11.841 | 90.526 | 8.926 | 9.106 | 8.873 |
| | BG/NBD | 124.992 | **7.725** | 4.096 | 11.516 | 131.352 | 10.809 | 7.560 | 11.573 |
| | PDO | 111.038 | 8.123 | 6.367 | 10.880 | **66.809** | 6.774 | 10.523 | 5.071 |
| | HB1 | **110.832** | 8.302 | 6.923 | 10.666 | 67.110 | 6.852 | 10.598 | 5.205 |
| | HB2 | 110.910 | 8.335 | 7.009 | 10.647 | 66.822 | **6.664** | 10.672 | 4.803 |
| | HB3 | 111.485 | 8.473 | 7.371 | 10.513 | 67.495 | 6.986 | 10.430 | 5.505 |
| | HB4 | 111.323 | 8.442 | 7.292 | 10.559 | 67.337 | 6.949 | 10.466 | 5.427 |
| CDNOW | Pareto/NBD | 4.789 | 1.282 | 0.886 | 2.411 | 114.655 | 8.899 | 6.353 | 14.758 |
| | BG/NBD | 4.788 | 1.276 | 0.879 | 2.377 | 114.640 | 8.942 | 6.462 | 14.647 |
| | PDO | **4.786** | 1.286 | 0.888 | 2.446 | **114.610** | 8.940 | 6.455 | 14.683 |
| | HB1 | 5.455 | 1.087 | 0.227 | 2.370 | 130.332 | **7.547** | 2.772 | 16.282 |
| | HB2 | 5.426 | 1.090 | 0.244 | 2.352 | 129.251 | 7.586 | 2.895 | 16.282 |
| | HB3 | 5.448 | **1.088** | 0.231 | 2.365 | 130.195 | 7.551 | 2.787 | 16.265 |
| | HB4 | 5.442 | 1.089 | 0.235 | 2.362 | 129.796 | 7.567 | 2.835 | 16.271 |
| grocer | Pareto/NBD | 83.958 | 5.454 | 3.554 | 11.381 | **719.044** | 24.024 | 19.359 | 31.472 |
| | BG/NBD | 84.097 | **5.341** | 3.342 | 11.503 | 720.197 | 24.341 | 20.457 | 30.755 |
| | PDO | **83.949** | 5.435 | 3.517 | 11.413 | 719.137 | 24.082 | 19.571 | 31.323 |
| | HB1 | 84.081 | 5.650 | 3.900 | 11.298 | 719.229 | **24.001** | 19.274 | 31.532 |

iterations 300,000 times, of which only the last 10,000 iterations were used[4]. For the BG/NBD and PDO models, we use a two-step Gibbs algorithm with 30,000 iterations, of which only the last 8,000 draws are used.

For metrics like the transaction frequency of a customer with history $(x_i, t_{x,i}, T_i)$, closed-form expressions for the Pareto/NBD, BG/NBD and PDO models are available conditional on both hyperparameters and behavioral parameters. This allows us to test our procedure based on the posterior draws on individual's parameters. We compare our simulation-based predictions to the results computed by the closed-form expressions conditioned on hyperparameters given in Schmittlein, Morrison, and Colombo (1987), Fader, Hardie, and Lee (2005a) and Jerath, Fader, and Hardie (2011). In all cases, the correlation between the expectations is more than 99.995%.

We consider the number of transactions in the out-of-sample period as well as the timing of the first out-of-sample transaction. More precisely, with the timing of the first out-of-sample transaction, we mean the minimum of the timing of the next transaction and the end of the out-of-sample period. We use MSE, MAE and the correlation between predicted and observed values. As the above measures do not distinguish between over- and underpredictions, we also provide the mean over all positive errors (ME+: overprediction) and the mean over all negative errors (ME−: underprediction).

[4]We use an extreme number of burn-in iterations, in practice convergence is achieved much earlier.

### 5.3.1 Predicting future transaction frequency

Table 10 summarizes the predictive performance on the number of future transactions. The HB models perform best in terms of the MSE, MAE and correlation measures on the grocer and the online grocer datasets. Taking into account that the covariate information works well for the online grocer, the HB2 model performs, consequently, the best among the HB models. For this model, the coefficient of the average number of items in the shopping basket is significant at the 5% level (based on the highest posterior density [HPD] interval). Adding the average spending worsens the out-of-sample predictions on transaction frequency. Therefore, the HB3 and HB4 models do not perform as well.

The good predictive performance of the HB model can be explained by the relaxation of the independence assumption in the heterogeneity distribution. Note that the HB and the Pareto/NBD models share the same individual-level assumptions. To further investigate the dependence, we take a look at the estimated correlations between purchase and defection rates. As emphasized by Abe (2009a), it makes most sense to look at the estimated correlations for the no-covariate configuration of the HB models (HB1 and HB3). Table 11 reports the posterior mean correlations for each pair of parameters on each dataset for the HB3 model, together with the highest posterior density regions (Hyndman 1996). We find a strong and significant negative correlation between purchase and defection rates for the online grocery data. Accordingly, we see a remarkable improvement on the prediction performance of the HB models on this dataset. We find a significant, but relatively smaller, negative correlation on the grocery data. The HB1 model performs only slightly better than the other models on this data. There is no significant correlation between the purchase and defection rates for the CDNOW dataset, and consequently, the Pareto/NBD model is the best predicting model with its more flexible gamma heterogeneity distribution.

The final two columns in Table 10 summarize the model's performance with regard to over- (ME+) and underpredictions (ME−). We find that for the Pareto/NBD model, the magnitude of underpredictions is bigger than that of overpredictions on all datasets. For the other models, the difference between ME+ and ME− depends on the data. The average underprediction is always larger than the average overpredictions on the CDNOW and grocery retailer datasets. It is exactly the other way around for the online grocer data, where the customers are relatively more loyal to the company. To further elaborate on this, we construct Table 12. This table presents summary statistics on the group of observations that are under- or overpredicted. We list the size of the group, mean values of the purchase frequency ($\overline{x}$) and the recency ($\overline{T} - \overline{t_x}$) in the calibration period, observed frequency in the holdout period ($\overline{x^*}$)

Table 10: Model's prediction performance on the number of transactions

| | | Correlation | MSE | MAE | ME+ | ME− |
|---|---|---|---|---|---|---|
| online grocer | Pareto/NBD | 0.9207 | 21.556 | 3.055 | 2.344 | 3.830 |
| | BG/NBD | 0.9195 | 20.840 | 2.996 | 3.253 | 2.340 |
| | PDO | 0.9169 | 21.219 | 3.047 | 3.347 | 2.343 |
| | HB1 | 0.9243 | 18.807 | 2.806 | 3.008 | 2.363 |
| | HB2 | **0.9250** | **18.543** | **2.779** | 2.941 | 2.419 |
| | HB3 | 0.9218 | 20.242 | 2.942 | 3.089 | 2.530 |
| | HB4 | 0.9221 | 20.168 | 2.934 | 3.075 | 2.538 |
| CDNOW | Pareto/NBD | **0.6304** | **2.568** | 0.754 | 0.429 | 1.866 |
| | BG/NBD | 0.6248 | 2.589 | 0.787 | 0.456 | 1.831 |
| | PDO | 0.6214 | 2.709 | 0.903 | 0.696 | 1.737 |
| | HB1 | 0.6235 | 2.962 | 0.717 | 0.209 | 2.083 |
| | HB2 | 0.6127 | 2.954 | 0.736 | 0.253 | 2.054 |
| | HB3 | 0.6241 | 2.743 | 0.680 | 0.234 | 2.090 |
| | HB4 | 0.6223 | 2.740 | **0.678** | 0.236 | 2.095 |
| grocer | Pareto/NBD | 0.8230 | 0.954 | 0.398 | 0.242 | 1.615 |
| | BG/NBD | 0.8216 | 0.966 | 0.416 | 0.265 | 1.602 |
| | PDO | 0.8189 | 0.983 | 0.460 | 0.317 | 1.591 |
| | HB1 | **0.8238** | **0.951** | **0.394** | 0.239 | 1.600 |

Note that ME+ and ME− give the average of over- and underpredictions over the groups

Table 11: 95% Highest Posterior Density Region and mean of correlations between behavioral rates

| | $\rho_{\theta_\lambda\theta_\mu}$ | | | $\rho_{\theta_\lambda\theta_\eta}$ | | | $\rho_{\theta_\eta\theta_\mu}$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | HPDR | | mean | HPDR | | mean | HPDR | | mean |
| online grocer | -0.718 | -0.297 | -0.501* | 0.694 | 0.770 | 0.732* | -0.765 | -0.687 | -0.730* |
| CDNOW | -0.215 | 0.197 | -0.011 | 0.235 | 0.421 | 0.332* | -0.729 | -0.675 | -0.703* |
| grocer | -0.259 | -0.115 | -0.184* | - | - | - | - | - | - |

* Indicates that 0 is not contained in the 95% HPDR (highest posterior density region).

and predictions ($\overline{\mathbb{E}[x]}$) for both groups. All models overpredict the transaction frequency, $x$, for the majority of customers in each datasets. In general, the overprediction occurs for those customers with a low transaction frequency and a long recency; and vice versa for the underprediction. In other words, the BTYD models overestimate transaction frequency for incidental buyers and underestimate it for frequent buyers.

We next study the relation between the prediction error and the number of in-sample purchases. The plots in Figure 1 show the average predicted number of out-of-sample purchases as a function of the number of in-sample purchases. Figure 2 gives the MAE as a function of the number of in-sample purchases. To be able to focus on the main differences between the model classes, we do not show the

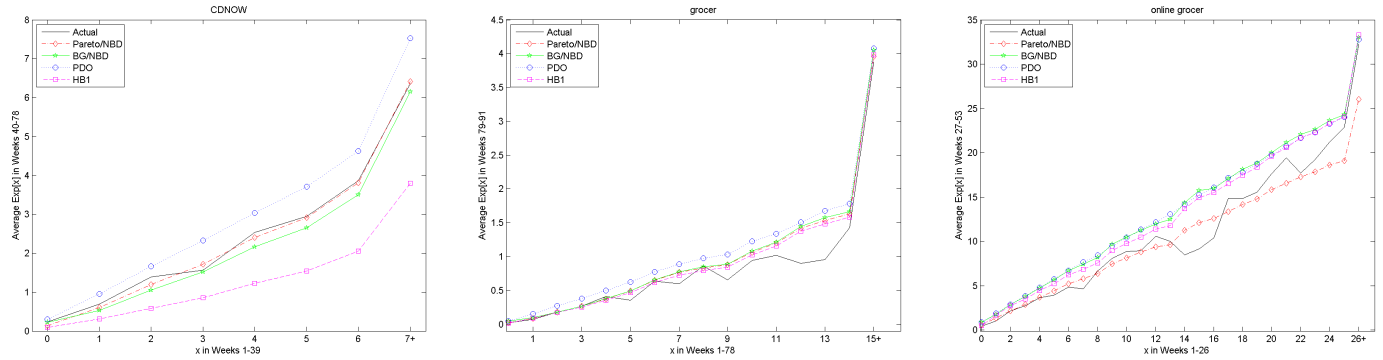
Table 12: Statistics on the groups of over- and underpredictions of future transaction frequency

| | | Overpredicted observations | | | | | | Underpredicted observations | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | ME+ | cus. % | $\overline{x}$ | $\overline{(T-t_x)}$ | $\overline{x^*}$ | $\overline{\mathbb{E}[x]}$ | ME− | cus. % | $\overline{x}$ | $\overline{(T-t_x)}$ | $\overline{x^*}$ | $\overline{\mathbb{E}[x]}$ |
| online grocer | Pareto/NBD | 2.344 | 52 | 6.593 | 8.855 | 3.138 | 5.482 | 3.830 | 48 | 13.984 | 3.705 | 14.934 | 11.104 |
| | BG/NBD | 3.253 | 72 | 8.912 | 6.887 | 6.072 | 9.325 | 2.340 | 28 | 13.243 | 5.119 | 15.710 | 13.371 |
| | PDO | 3.347 | 70 | 8.730 | 7.030 | 5.795 | 9.142 | 2.343 | 30 | 13.412 | 4.889 | 15.787 | 13.444 |
| | HB1 | 3.008 | 69 | 8.909 | 6.998 | 5.917 | 8.925 | 2.363 | 31 | 12.806 | 5.058 | 15.061 | 12.698 |
| | HB2 | 2.941 | 69 | 8.961 | 7.037 | 5.993 | 8.934 | 2.419 | 31 | 12.733 | 4.949 | 14.993 | 12.574 |
| | HB3 | 3.089 | 74 | 8.908 | 6.599 | 6.172 | 9.261 | 2.530 | 26 | 13.560 | 5.802 | 16.109 | 13.580 |
| | HB4 | 3.075 | 74 | 8.944 | 6.573 | 6.212 | 9.287 | 2.538 | 26 | 13.482 | 5.872 | 16.047 | 13.509 |
| CDNOW | Pareto/NBD | 0.429 | 77 | 0.851 | 27.303 | 0.170 | 0.598 | 1.866 | 23 | 1.695 | 20.977 | 2.946 | 1.079 |
| | BG/NBD | 0.456 | 76 | 0.813 | 27.698 | 0.144 | 0.600 | 1.831 | 24 | 1.764 | 20.113 | 2.859 | 1.028 |
| | PDO | 0.696 | 80 | 0.913 | 26.942 | 0.216 | 0.912 | 1.737 | 20 | 1.564 | 21.567 | 3.136 | 1.399 |
| | HB1 | 0.209 | 73 | 0.631 | 28.748 | 0.041 | 0.250 | 2.083 | 27 | 2.116 | 18.666 | 2.836 | 0.753 |
| | HB2 | 0.253 | 73 | 0.639 | 28.672 | 0.046 | 0.299 | 2.054 | 27 | 2.111 | 18.760 | 2.853 | 0.798 |
| | HB3 | 0.234 | 76 | 0.733 | 27.811 | 0.108 | 0.342 | 2.090 | 24 | 1.982 | 20.343 | 2.977 | 0.887 |
| | HB4 | 0.236 | 76 | 0.742 | 27.744 | 0.115 | 0.351 | 2.095 | 24 | 1.968 | 20.477 | 2.988 | 0.893 |
| grocer | Pareto/NBD | 0.242 | 89 | 3.516 | 51.082 | 0.145 | 0.387 | 1.615 | 11 | 12.464 | 17.209 | 3.533 | 1.918 |
| | BG/NBD | 0.265 | 89 | 3.573 | 51.029 | 0.155 | 0.420 | 1.602 | 11 | 12.105 | 17.298 | 3.489 | 1.887 |
| | PDO | 0.317 | 89 | 3.541 | 51.002 | 0.149 | 0.466 | 1.591 | 11 | 12.411 | 17.287 | 3.561 | 1.970 |
| | HB1 | 0.239 | 88 | 3.404 | 51.180 | 0.152 | 0.391 | 1.600 | 12 | 12.095 | 17.151 | 3.450 | 1.850 |

results for the HB models including spending and/or covariates.

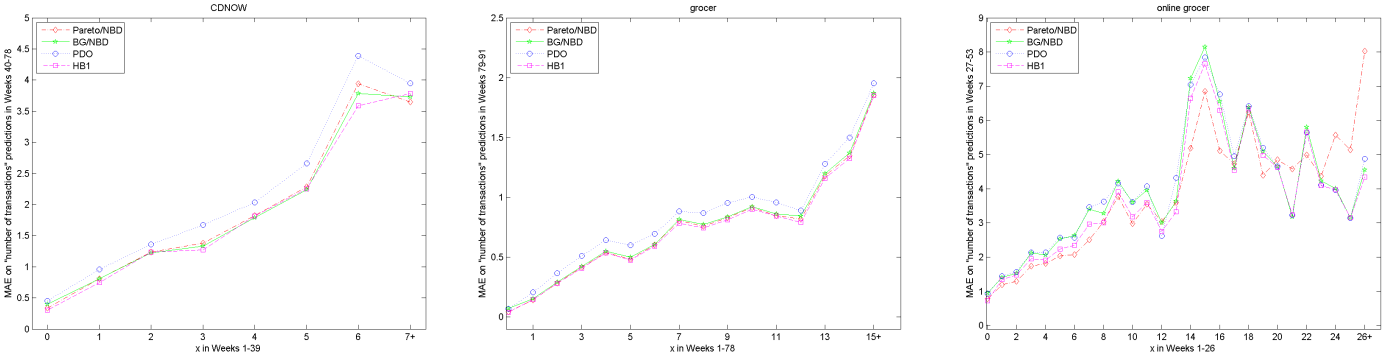
The PDO model tends to yield higher predictions for CDNOW data. This matches our findings in Tables 10 and 12. On average, the HB1 model yields the lowest predicted transaction numbers. Remarkably, this is not reflected in a poor forecasting performance for this model. In fact, Figure 2a shows that the HB1 model predicts very well for all values of the in-sample number of transactions. For the grocer dataset, all models show a very similar prediction pattern. Only the PDO model stands out with its relatively high predictions. Figure 2b shows that this leads to higher MAEs. The Pareto/NBD model is different from the other models for the online grocer data. This model has the tendency to underpredict transaction numbers (see also Tables 10 and 12).

The MAE tends to increase with the number of in-sample transaction numbers for the CDNOW and grocer datasets, contrasting with what is observed for the online grocery data (see Figure 2). The online grocer dataset stands out with its data center leaning toward frequent buyers. The predictions now result from models pulling values to this center.




(a) (b) (c)

Figure 1: Conditional expectation of future transaction numbers on CDNOW, grocer and online grocer datasets. All plots right-censor the horizontal axis for readability. For CDNOW data, the group having $\geq 7$ repeat-purchases corresponds to only 3% of the observations; for the grocer dataset 9% of the observations are in the group $\geq 15$; and for the online grocer 6% are $\geq 26$.




(a) (b) (c)

Figure 2: MAE on the number of future transaction predictions on CDNOW, grocer and online grocer datasets

### 5.3.2 Predicting future transaction timing

Finally, we consider the performance on predicting future transaction timing[5]. Table 13 presents an overview of the main results. Interestingly, the PDO model has a good performance on the CDNOW and grocer datasets. This model did not perform particularly well on predicting the number of transactions. Note that the timing of transactions is strongly influenced by the defection process and that the PDO model specially focuses on this process. Jerath, Fader, and Hardie (2011) demonstrate that the PDO model allows the defection process to be somewhere in between the extremes implied by the Pareto/NBD model and the no-defection NBD model. The PDO model performs the worst on the

[5]We thank Batislam, Denizel, and Filiztekin (2007) and Fader, Hardie, and Lee (2005b) for making the out-of-sample timing data available.

online grocer data. One reason may be the long (estimated) defection period interval ($\tau = 20.001$ days).

The HB models also perform rather well on the grocer and online grocer datasets. For both datasets we found a significant correlation between the behavioral parameters. Among the HB models, a remarkable point is the improved performance of the HB3 model when taking into account the average spending amount on CDNOW and online grocer datasets. This can be explained by the existence of the strong and significant negative correlation between the spending and defection parameters in both datasets (see Table 11).

Table 13: Model's prediction performance on the timing of next transaction

| | | Correlation | MSE | MAE | ME+ | ME− |
|---|---|---|---|---|---|---|
| online grocer | Pareto/NBD | 0.7296 | 46.674 | 4.508 | 2.649 | 5.801 |
| | BG/NBD | 0.7259 | 47.173 | 4.523 | 2.668 | 5.792 |
| | PDO | 0.6780 | 50.668 | 5.116 | 3.152 | 7.769 |
| | HB1 | **0.7328** | **43.416** | 4.223 | 2.991 | 5.134 |
| | HB2 | 0.7254 | 44.374 | 4.296 | 3.068 | 5.210 |
| | HB3 | 0.7201 | 46.594 | **4.067** | 2.973 | 4.772 |
| | HB4 | 0.7204 | 46.504 | 4.073 | 2.983 | 4.777 |
| CDNOW | Pareto/NBD | 0.5789 | 125.451 | **7.372** | 17.013 | 4.027 |
| | BG/NBD | 0.5750 | 125.153 | 8.122 | 17.027 | 5.033 |
| | PDO | **0.5828** | **123.441** | 8.517 | 15.343 | 6.228 |
| | HB1 | 0.5486 | 273.555 | 15.660 | 10.062 | 17.051 |
| | HB2 | 0.5449 | 282.423 | 15.865 | 9.781 | 17.352 |
| | HB3 | 0.5687 | 270.514 | 15.408 | 9.229 | 16.898 |
| | HB4 | 0.5689 | 270.028 | 15.376 | 9.214 | 16.850 |
| grocer | Pareto/NBD | 0.8183 | 7.684 | 1.442 | 4.590 | 1.182 |
| | BG/NBD | 0.8192 | 7.770 | 1.542 | 4.551 | 1.293 |
| | PDO | **0.8226** | 7.976 | 1.734 | 4.469 | 1.514 |
| | HB1 | 0.8190 | **7.602** | **1.426** | 4.639 | 1.171 |

ME+ and ME− give the average over the groups of overpredictions and underpredictions

In Table 14, we investigate for what type of observation the purchase time is over- or underpredicted. We present the size of the over- and underpredicted group, group-specific characteristics in the calibration period, the average observed timing ($\overline{t_f^*}$) in the holdout period and the average predicted time ($\overline{\mathbb{E}[t_f]}$). In line with the previous results, all BTYD models underpredict the timing of the next purchase for customers who have a low transaction frequency and high recency; and vice versa for the groups of higher predictions.

In Figure 3, we show the average predictions as a function of the time of the last in-sample transaction ($t_x$). Note that the timing predictions are explicitly influenced by $t_x$ (see Equations (4), (8), and (12)).

Table 14: Statistics on the groups of over- and underpredictions of future transactions timing

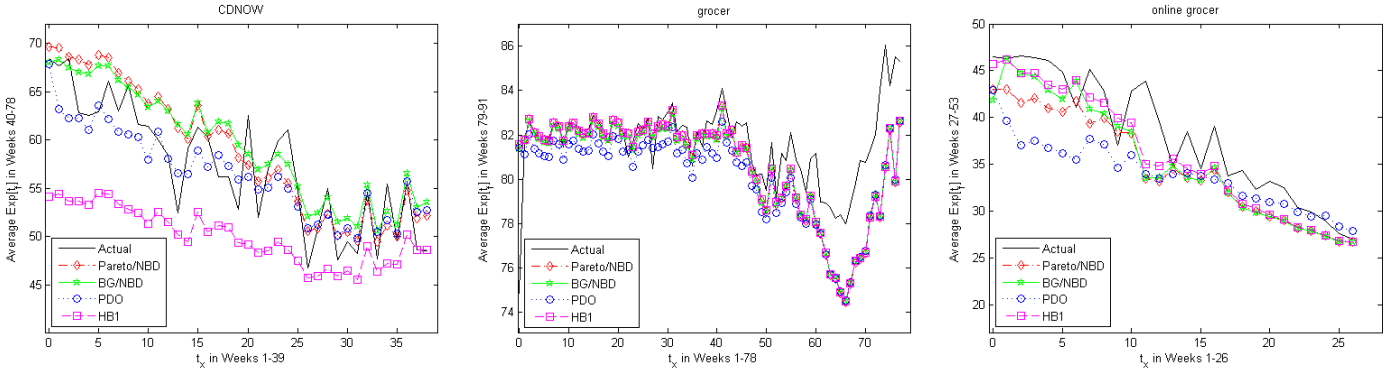
| | | Overpredicted observations | | | | | | Underpredicted observations | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | ME+ | cus. % | $\overline{x}$ | $\overline{(T-t_x)}$ | $\overline{t_f^*}$ | $\overline{\mathbb{E}[t_f]}$ | ME− | cus. % | $\overline{x}$ | $\overline{(T-t_x)}$ | $\overline{t_f^*}$ | $\overline{\mathbb{E}[t_f]}$ |
| online grocer | Pareto/NBD | 2.65 | 41 | 11.80 | 4.47 | 27.49 | 30.14 | 5.80 | 59 | 8.97 | 7.72 | 37.93 | 32.13 |
| | BG/NBD | 2.67 | 41 | 11.79 | 4.49 | 27.51 | 30.18 | 5.79 | 59 | 8.99 | 7.69 | 37.85 | 32.05 |
| | PDO | 3.15 | 57 | 13.71 | 3.38 | 27.39 | 30.54 | 7.77 | 43 | 5.30 | 10.45 | 42.11 | 34.34 |
| | HB1 | 2.99 | 43 | 11.69 | 4.61 | 27.60 | 30.59 | 5.13 | 57 | 8.97 | 7.71 | 38.13 | 32.99 |
| | HB2 | 3.07 | 43 | 11.62 | 4.57 | 27.60 | 30.67 | 5.21 | 57 | 9.03 | 7.75 | 38.15 | 32.94 |
| | HB3 | 2.97 | 39 | 11.73 | 4.86 | 27.59 | 30.56 | 4.77 | 61 | 9.10 | 7.38 | 37.55 | 32.78 |
| | HB4 | 2.98 | 39 | 11.75 | 4.85 | 27.58 | 30.56 | 4.78 | 61 | 9.09 | 7.38 | 37.57 | 32.79 |
| CDNOW | Pareto/NBD | 17.01 | 26 | 2.26 | 18.52 | 43.49 | 60.51 | 4.03 | 74 | 0.62 | 28.42 | 71.34 | 66.31 |
| | BG/NBD | 17.03 | 26 | 2.21 | 18.80 | 43.68 | 60.69 | 5.03 | 74 | 0.64 | 28.32 | 71.28 | 67.25 |
| | PDO | 15.34 | 25 | 2.27 | 18.46 | 43.04 | 58.39 | 6.23 | 75 | 0.63 | 28.36 | 71.25 | 65.03 |
| | HB1 | 10.06 | 20 | 2.45 | 17.60 | 39.80 | 49.86 | 17.05 | 80 | 0.69 | 28.10 | 70.22 | 53.17 |
| | HB2 | 9.78 | 20 | 2.50 | 17.40 | 39.62 | 49.40 | 17.35 | 80 | 0.69 | 28.12 | 70.17 | 52.82 |
| | HB3 | 9.23 | 19 | 2.38 | 18.29 | 39.80 | 49.03 | 16.90 | 81 | 0.72 | 27.88 | 70.05 | 53.15 |
| | HB4 | 9.21 | 19 | 2.38 | 18.33 | 39.74 | 48.96 | 16.85 | 81 | 0.72 | 27.85 | 70.01 | 53.16 |
| grocer | Pareto/NBD | 4.59 | 8 | 7.17 | 23.32 | 75.03 | 79.62 | 1.18 | 92 | 4.32 | 49.21 | 82.77 | 81.59 |
| | BG/NBD | 4.55 | 8 | 7.10 | 23.21 | 75.05 | 79.60 | 1.29 | 92 | 4.32 | 49.22 | 82.77 | 81.48 |
| | PDO | 4.47 | 7 | 7.18 | 22.88 | 74.81 | 79.28 | 1.51 | 93 | 4.32 | 49.18 | 82.77 | 81.26 |
| | HB1 | 4.64 | 8 | 7.18 | 23.30 | 75.03 | 79.67 | 1.17 | 91 | 4.18 | 49.23 | 82.76 | 81.59 |

We show the corresponding MAE values in Figure 4. Figure 3a clearly shows that the HB1 model gives quite different predictions compared to the other models for CDNOW; for HB1 the predictions tend to be smaller. Based on Figure 4a we conclude that these predictions are too low. The MAE for the HB1 model is the highest among all models. However, for the recent buyers (high $t_x$ values) the differences between the models are relatively small.

For the grocer dataset, we see that all the models, except the PDO model, have almost identical predictions and performance for the non-recent buyers (see Figures 3b and 4b). The PDO model has lower predictions and higher MAE for those customers. Again for recent buyers, all models have very similar predictions so that it is difficult to distinguish between the models for this group of observations.

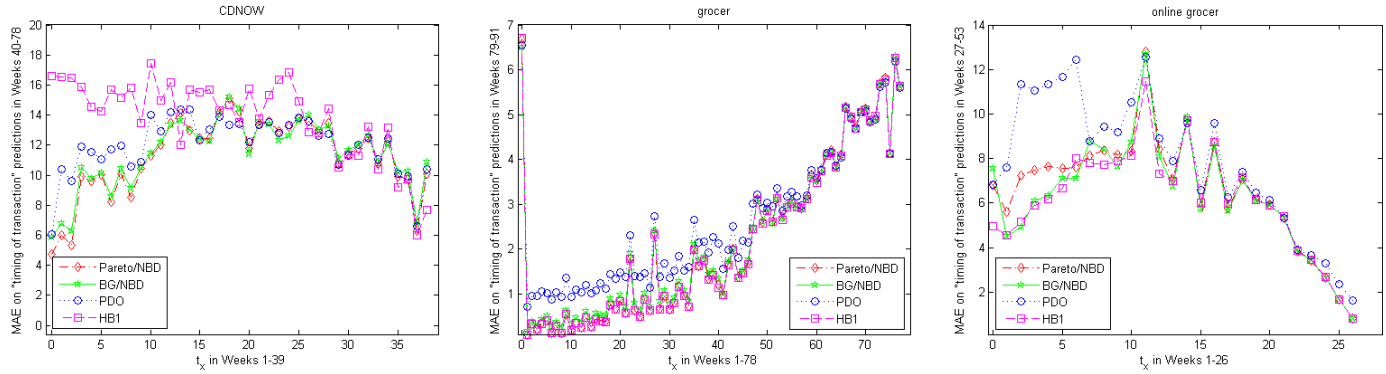
For the online grocer data, the PDO model also performs relatively poorly for non-recent buyers (see Figures 3c and 4c). The PDO model tends to underpredict the timing of the first transaction for customers who do not have recent transactions. On this data, the majority of customers are frequent buyers who had recent transactions. For instance, the percentage of customers who have $t_x \leq 10$ weeks is just 15% and therefore the left hand side of the figure does not have a big weight in the overall predictive performance of the models for this dataset. However, for the other datasets, a large part of the dataset have low values of $t_x$ (53% of customers have $t_x \leq 10$ on the grocery dataset and 73% of

customers has $t_x \leq 10$ on the CDNOW dataset).




(a) (b) (c)

Figure 3: Conditional expectation of future transaction timing on CDNOW, grocer and online grocer datasets




(a) (b) (c)

Figure 4: MAE of future transaction timing predictions on CDNOW, grocer and online grocer datasets

## 6 Discussion

In this paper, our aim is to present a new use of the existing buy-till-you-defect [BTYD] models. In the current literature, the main focus is on predicting the transaction frequency. We argue that prediction of the future transaction timing of an individual is also very relevant. For each of the most popular BTYD models, we develop a method to calculate such predictions.

First of all, these timing predictions are useful to compare the quality of the existing models on an additional metric. Next, timing predictions have a clear managerial purpose. For example, consider an online retailer implementing micro-marketing strategies. The most appropriate time to contact its

customers depends on their expected timing of the next purchase. High quality timing predictions may contribute to achieving the full potential of micro-marketing (Zhang and Krishnamurthi 2004).

Following the pioneering research by Gupta (1988), there is a growing literature that examines the effectiveness of promotions on whether to buy, '*when*' to buy, and how much to buy (see the summary of relevant literature in Gönül and Hofstede (2006)). We believe that using the BTYD models to predict the timing of transactions provides a new means of answering the 'when' question.

An operations manager may also use predictions on the timing and transaction value as input for Revenue Management. For example, online retailers have limited delivery capacity at a given time. Given the appropriate predictions, operations managers can prioritize valued customers for highly demanded delivery time slots (Talluri and Van Ryzin 2005). Tereyağoğlu, Fader, and Veeraraghavan (2012) emphasize the crucial role of having accurate timing predictions to improve revenues. In summary, we believe that the ability to predict the timing of future transactions can be helpful to accelerate research on aforementioned topics in industries that operate in a noncontractual setting.

We present a general method and specific formulas that can be used to predict the timing of the next purchase for four of the established BTYD models. Such formulas have not been presented before. We use these methods to compare the predictive performance of all models on three very different datasets. We find that the predictive performance of the models varies not only with the characteristics of the data, but also with respect to the performance metric.

Managers who aim to forecast their customers' transaction frequency should first examine general characteristics of the customer cohort and then choose the best fitting model. The HB models tend to perform relatively poorly in case data is weak due to many zero-repeat buyers. On the other hand, they do have a clear advantage if there are many repeat buyers and there are significant correlations between the behavioral parameters.

The PDO and HB models perform well on the timing of transaction predictions, again conditional on some data characteristics. Our conclusions on model choice are based on informally relating data characteristics to forecasting performance on just three datasets. There are studies that attempt to formally quantify and validate such relations through classification and regression trees and random forests (Schwartz, Bradlow, and Fader 2012). Such a formal study is very welcome in this context to arrive at more general recommendations.

By comparing the predictive performance on future frequency versus timing, we found that the BTYD models perform rather poorly on the latter. A closer focus on the defection process may lead

to better timing predictions. The ideas of Bueschken and Ma (2012) may be helpful in this context. They provide a new perspective on possible switches between active and inactive states, and allow for both regular and incidental buyers by relaxing the Poisson process assumption on the arrival of transactions.

# Appendix A Timing expressions

In this section, we present the derivations of the expected timing of the last transaction, $t_x$, in the observation period $[0,T]$ and the expected timing of the next event (either the first purchase or the end of the forecast interval), $t_f$, conditioned on an individual's parameters. The hyperparameters do not play a role here. In all sections of this appendix we drop the $i$ subscript, representing customer $i$, for notational simplicity. In the notation we also do not condition on the length of the observational interval $T$.

## Appendix A.1 Timing of transactions for Pareto/NBD and HB models

The derivations in this section apply to the original Pareto/NBD model and its HB extension. The expressions are the same as both models have the same assumptions on individual behavior. The time of defection, $t_\Delta$, has the probability function[6]

$$\mathbb{P}(\mathrm{d}t_\Delta|\lambda,\mu) = \mu e^{-\mu t_\Delta}\,\mathrm{d}t_\Delta\,. \tag{15}$$

Setting $t_\delta = \min(t_\Delta, T)$, we obtain

$$\mathbb{P}(\mathrm{d}t_\delta|\lambda,\mu) = \begin{cases} \mu e^{-\mu t_\delta}\,\mathrm{d}t_\delta & \text{if } 0 \le t_\delta < T \\ e^{-\mu T}\delta_T(t_\delta)\mathrm{d}t_\delta & \text{if } t_\delta = T \\ 0 & \text{otherwise,} \end{cases} \tag{16}$$

where $\delta_w(x)$ is the Dirac-delta function at $w$ evaluated at $x$[7]. Conditioning on the unobserved value $t_\delta$, we find the density of $t_x$ on $(0,T]$ as

$$\mathbb{P}(\mathrm{d}t_x|t_\delta,\lambda,\mu) = \left(\lambda e^{-\lambda(t_\delta - t_x)} + \delta_0(t_x)e^{-\lambda t_\delta}\right)\mathrm{d}t_x, \tag{17}$$

[6] We use a rather formal notation here as our stochastic variables have a mixed discrete/continuous distribution. For practical purposes one can see the part before $\mathrm{d}t_\Delta$ on the right-hand side of (15) as the traditional probability density function.

[7] More precisely, $\delta_w()$ is a point mass at $w$ normalized such that for any continuous function $g$, $\int g(t)\delta_w(t)\,\mathrm{d}t = g(w)$.

where we make use of the memoryless property of the Poisson process. Informally, we can look back in time and do as if the process starts at $t_\delta$. Integrating over $t_\delta$, one obtains

$$\mathbb{P}(\mathrm{d}t_x|\lambda,\mu)=\int_{t_\delta\in[t_x,T]}\mathbb{P}(\mathrm{d}t_x|t_\delta,\lambda,\mu)\mathbb{P}(\mathrm{d}t_\delta|\lambda,\mu)=\begin{cases}\lambda\,\frac{\mu e^{-(\lambda+\mu)t_x}+\lambda e^{-(\lambda+\mu)T}}{\lambda+\mu}\,\mathrm{d}t_x & \text{if } 0<t_x\le T\\ \left(\frac{\mu}{\lambda+\mu}+\frac{\lambda e^{-(\lambda+\mu)T}}{\lambda+\mu}\right)\delta_0(t_x)\,\mathrm{d}t_x & \text{if } t_x=0\,.\end{cases} \tag{18}$$

Based on Equation (18), the expected value on the time of the last transaction is calculated as follows,

$$\mathbb{E}(t_x|\lambda,\mu)=\int_0^\infty t_x\,\mathbb{P}(\mathrm{d}t_x|\lambda,\mu)=\frac{1-e^{-\mu T}}{\mu}-\frac{1-e^{-(\lambda+\mu)T}}{\lambda+\mu}\,. \tag{19}$$

Next, we present the derivations for the predictions of the time of next event from the end of the calibration period conditional on $x$ and $t_x$: $\mathbb{E}(t_f|x,t_x,\lambda,\mu)$. Let $T^+$ be some future horizon $T^+>T$. Consider the first future transaction after $T$. We define $t_f$ as the time of this occurrence or $T^+$, whichever is first. We have

$$\mathbb{E}(t_f|x,t_x,\lambda,\mu)=\mathbb{E}(t_f|x,t_x,z=1,\lambda,\mu)\,p^+ +\mathbb{E}(t_f|x,t_x,z=0,\lambda,\mu)(1-p^+),$$

where $z=1$ indicates that a customer is active at time $T$ and

$$p^+=\mathbb{E}(z|x,t_x,\lambda,\mu)=\frac{\lambda}{\lambda+e^{(\lambda+\mu)(T-t_x)}}\,. \tag{20}$$

Consider an active customer. Then the density of the first timing, $t$, of a transaction on $(T,\infty)$ is $\lambda e^{-(\lambda+\mu)(t-T)}$ and $t$ has a point mass at infinity of $\frac{\mu}{\lambda+\mu}$ as defection may have been the first event to happen. Therefore, on the interval $(T,T^+]$ the density of $t_f$ given a customer's transaction data and that the customer is active at time $T$ is $\pi_f(t|x,t_x,z=1,\lambda,\mu)=\lambda e^{-(\lambda+\mu)(t-T)}$. The expectation is computed as,

$$\begin{aligned}\mathbb{E}(t_f|x,t_x,\lambda,\mu)&=p^+\int_T^{T^+}t\,\pi_f(t|x,t_x,z=1,\lambda,\mu)\,\mathrm{d}t\\&\quad+p^+\left(1-\int_T^{T^+}\pi_f(t|x,t_x,z=1,\lambda,\mu)\,\mathrm{d}t\right)T^+ +(1-p^+)T^+\\&=T+\frac{\mu e^{(\lambda+\mu)(T-t_x)}}{\lambda+\mu e^{(\lambda+\mu)(T-t_x)}}(T^+-T)+\frac{\lambda}{\lambda+\mu e^{(\lambda+\mu)(T-t_x)}}\,\frac{1-e^{-(\lambda+\mu)(T^+-T)}}{\lambda+\mu}\end{aligned} \tag{21}$$

## Appendix A.2 Timing of transactions for BG/NBD model

In the BG/NBD model, the timing of defection, $t_\Delta$, is also the timing of the last transaction and its density is

$$\mathbb{P}(\mathrm{d}t_\Delta|\lambda,p)=\lambda p e^{-\lambda p t_\Delta}\,\mathrm{d}t_\Delta, \tag{22}$$

see Fader, Hardie, and Lee 2005a. It should be noted that the first purchase at time 0 is special in that a customer cannot defect at time 0. Given that $t_\delta = \min(t_\Delta, T)$:

$$\mathbb{P}(\mathrm{d}t_\delta|\lambda,p) = \begin{cases} \lambda p e^{-\lambda p t_\delta}\,\mathrm{d}t_\delta & \text{if } 0 < t_\delta < T \\ e^{-\lambda p T}\delta_T(t_\delta)\,\mathrm{d}t_\delta & \text{if } t_\delta = T \\ 0 & \text{otherwise}. \end{cases} \tag{23}$$

Conditioning on the unobserved value $t_\delta$, we find the density of $t_x$ as

$$\mathbb{P}(\mathrm{d}t_x|t_\delta,\lambda,p) = \begin{cases} \delta_{t_\delta}(t_x)\,\mathrm{d}t_x & \text{if } t_\delta < T \\ \left(\lambda(1-p)e^{-\lambda(1-p)(T-t_x)} + e^{-\lambda(1-p)T}\delta_0(t_x)\right)\mathrm{d}t_x & \text{if } t_\delta = T \\ 0 & \text{otherwise}. \end{cases} \tag{24}$$

Integrating over $t_\delta$, one obtains the probability

$$\begin{aligned} \mathbb{P}(\mathrm{d}t_x|\lambda,p) &= \int_{t_\delta\in[t_x,T]} \mathbb{P}(\mathrm{d}t_x|t_\delta,\lambda,p)\mathbb{P}(\mathrm{d}t_\delta|\lambda,p) \\ &= \left(\lambda p e^{-\lambda p t_x} + (1-p)\lambda e^{-\lambda(T-(1-p)t_x)} + e^{-\lambda T}\delta_0(t_x)\right)\mathrm{d}t_x\,. \end{aligned}$$

and, therefore

$$\mathbb{P}(\mathrm{d}t_x|\lambda,p) = \begin{cases} \lambda\left(p e^{-\lambda p t_x} + (1-p)e^{-\lambda T}e^{\lambda(1-p)t_x}\right)\mathrm{d}t_x & \text{if } 0 < t_x \le T \\ e^{-\lambda T}\delta_0(t_x)\,\mathrm{d}t_x & \text{if } t_x = 0 \\ 0 & \text{otherwise}. \end{cases} \tag{25}$$

Using equation (25), the expected value of the time of the last transaction in the observation interval $[0, T]$ can be calculated as

$$\mathbb{E}(t_x|\lambda,p) = \int_0^T t_x\lambda\left(p e^{-\lambda p t_x} + (1-p)e^{-\lambda T}e^{\lambda(1-p)t_x}\right)\mathrm{d}t_x = \frac{1}{1-p}\left(\frac{1-e^{-\lambda p T}}{\lambda p} - \frac{1-e^{-\lambda T}}{\lambda}\right). \tag{26}$$

For the case $x, t_x > 0$ one easily sees, by referring to the Pareto/NBD result on $p^+$ in Equation (20) under substituting $(1-p)\lambda$ for $\lambda$ and $\lambda p$ for $\mu$, that

$$p^+ = \mathbb{P}(z=1|x,t_x,\lambda,p) = \begin{cases} \frac{1-p}{1-p+pe^{\lambda(T-t_x)}} & \text{if } x, t_x > 0 \\ 1 & \text{if } x = 0 = t_x\,. \end{cases} \tag{27}$$

The density of the first future transaction given the rates, the observed transaction data and the customer being active at $T$ is $\pi_f(t|x,t_x,z=1,\lambda,\mu) = \lambda e^{-\lambda(t-T)}$. Note that an active customer will always make at least one future

purchase. The expected value of the first future purchase timing (or $T^+$) is

$$\mathbb{E}(t_f|x,t_x,\lambda,p) = p^+\lambda \int_{t_f=T}^{t_f=T^+} t_f e^{-\lambda(t_f-T)}\,\mathrm{d}t_f + \left(1-p^+ + p^+ e^{-\lambda(T^+-T)}\right) T^+ \tag{28}$$

$$= T + (1-p^+)(T^+-T) + p^+ \frac{1-e^{-\lambda(T^+-T)}}{\lambda}\,. \tag{29}$$

## Appendix A.3 Timing of transactions for PDO model

In the periodic-defection-model (PDO) (Jerath, Fader, and Hardie 2011) the time of defection, $t_\Delta$, has a discrete distribution with support $\{n\tau\}_{n=1,2\ldots}$ which is given as

$$\mathbb{P}(t_\Delta = n\tau|\lambda,p) = p(1-p)^{n-1}, \tag{30}$$

where $\tau$ can be treated as a known value (estimated using MLE at the customer base level). Let $t_\delta = \min(t_\Delta, T)$ be the time after which no transactions are observed. Given $t_\delta$ the distribution of the time, $t_x$, of the last observed transaction in $[0,T]$ is

$$\mathbb{P}(\mathrm{d}t_x|t_\delta,\lambda,p) = I_{[0,t_\delta]}(t_x) e^{-\lambda(t_\delta-t_x)}(\lambda+\delta_0(t_x))\,\mathrm{d}t_x, \tag{31}$$

$I_A$ is the indicator function of the set $A$. Note the distribution's point mass at 0. One computes

$$\begin{aligned}\mathbb{P}(\mathrm{d}t_x|\lambda,p) &= \int_{t_\delta\in[t_x,T]} \mathbb{P}(\mathrm{d}t_x|t_\delta,\lambda,p)\mathbb{P}(\mathrm{d}t_\delta|\lambda,p) \\ &= \left(\sum_{n=m_x}^{N} p(1-p)^{n-1}e^{-\lambda n\tau} + (1-p)^N e^{-\lambda T}\right)(\lambda+\delta_0(t_x))\,e^{\lambda t_x}\,\mathrm{d}t_x\,. \end{aligned} \tag{32}$$

where we use the notations $N$ for $\lfloor T/\tau \rfloor$ and $m_x$ as the time of the first opportunity to defect after or at $t_x$, expressed as a multiple of $\tau$, that is, $m_x = \left\lfloor \frac{t_x}{\tau} + 1 \right\rfloor$. Using (32) together with the observation that in our case, it holds that

$$\sum_{m=1}^{N} \int_{\{m_x=m\}} \sum_{n=m}^{N} (\cdot)\,\mathrm{d}t_x = \sum_{n=1}^{N}\sum_{m=1}^{n} \int_{\{m_x=m\}} (\cdot)\,\mathrm{d}t_x = \sum_{n=1}^{N} \int_{t_x=0}^{t_x=n\tau} (\cdot)\,\mathrm{d}t_x\,,$$

the expected value for the time of the last observed transaction in the interval $[0,T]$ is found as[8]

$$\mathbb{E}(t_x|\lambda,p) = \sum_{n=1}^{N} p(1-p)^{n-1}\left(n\tau - \frac{1-e^{-n\lambda\tau}}{\lambda}\right) + (1-p)^N\left(T - \frac{1-e^{-\lambda T}}{\lambda}\right). \tag{33}$$

Now let us turn to the timing of the first repeat transaction, $t_1$, where, by convention, we set $t_1 = \infty$ in case there is no repeat transaction after the initial transaction at time 0. More in particular, we study $t_1$ capped by the

[8] For reasons of computational efficiency, in cases where $N$ is a large number, the summation in Equation (33) may be written as $\frac{\tau}{p}\left(N(1-p)^{(N+1)} - (N+1)(1-p)^N + 1\right) - \frac{1-(1-p)^N}{\lambda} + \frac{pe^{-\lambda\tau}}{\lambda}\frac{\left((1-p)e^{-\lambda\tau}\right)^N - 1}{(1-p)e^{-\lambda\tau}-1}$.

observation period's length, $t^+ = \min(t_1, T)$. Then, by analogy to (31) and (32) we obtain

$$\mathbb{P}(\mathrm{d}t^+|t_\delta,\lambda,p) = \left(I_{[0,t_\delta]}(t^+)\lambda e^{-\lambda t^+} + e^{-\lambda t_\delta}\delta_T(t^+)\right)\mathrm{d}t^+ \tag{34}$$

and

$$\begin{aligned}\mathbb{P}(\mathrm{d}t^+|\lambda,p) = &\left(\sum_{n=\lceil t^+/\tau\rceil}^{N} p(1-p)^{n-1}\lambda e^{-\lambda t^+} + (1-p)^N\right)\lambda e^{-\lambda t^+}\,\mathrm{d}t^+ \\ &+ \left(\sum_{n=1}^{N} p(1-p)^{n-1}e^{-n\lambda\tau} + (1-p)^N e^{-\lambda T}\right)\delta_T(t^+)\mathrm{d}t^+ .\end{aligned} \tag{35}$$

From the density in (35), the expected value for the timing of the first transaction becomes

$$\begin{aligned}\mathbb{E}(t^+|\lambda,p,T) = &\sum_{n=1}^{N} p(1-p)^{n-1}\left(\frac{1-(n\lambda\tau+1)e^{-n\lambda\tau}}{\lambda}\right) + (1-p)^N\left(\frac{1-(\lambda T+1)e^{-\lambda T}}{\lambda}\right) \\ &+ \left(p e^{-\lambda\tau}\frac{1-\left((1-p)e^{-\lambda\tau}\right)^N}{1-(1-p)e^{-\lambda\tau}} + (1-p)^N e^{-\lambda T}\right)T\end{aligned}$$

or

$$\begin{aligned}\mathbb{E}(t^+|\lambda,p,T) = 1/\lambda &\left(1 - \sum_{n=1}^{\lfloor T/\tau\rfloor} p(1-p)^{n-1}(n\lambda\tau+1)e^{-n\lambda\tau} - (1-p)^{\lfloor T/\tau\rfloor}(\lambda T+1)e^{-\lambda T}\right) \\ &+ \left(p e^{-\lambda\tau}\frac{1-\left((1-p)e^{-\lambda\tau}\right)^{\lfloor T/\tau\rfloor}}{1-(1-p)e^{-\lambda\tau}} + (1-p)^{\lfloor T/\tau\rfloor}e^{-\lambda T}\right)T .\end{aligned} \tag{36}$$

This expression for the timing of the first transaction in the calibration period is reused for calculating the timing of the first future transaction after $T$, see Equation (12).

# Appendix B Estimation procedure for Pareto/NBD, BG/NBD and PDO models

To calculate the various expectations, we also need draws from the conditional density of the individual-level parameters. Below we discuss how to obtain such draws for the Pareto/NBD, BG/NBD and PDO model.

For the BG/NBD and PDO models, the relevant parameters are the transaction rate, $\lambda$, and the probability of defection, $p$, per defection opportunity. Below, we argue that we can easily draw from the full conditional distributions $\pi(\lambda|x,t_x,p)$ and $\pi(p|x,t_x,\lambda)$. We rely on Gibbs sampling to obtain draws from the joint conditional distribution $\pi(\lambda,p|x,t_x)$.

For the Pareto/NBD model, sampling from the full conditionals is not straightforward. Therefore, we need to develop a different method. We propose to use a random-walk Metropolis-Hastings algorithm to obtain draws

from the individual-level posterior distribution.

## Appendix B.1 The Pareto/NBD model

The likelihood function for the Pareto/NBD model is

$$f(x,t_x|\lambda,\mu)=\frac{\lambda^x}{\lambda+\mu}(\mu e^{-(\lambda+\mu)t_x}+\lambda e^{-(\lambda+\mu)T})\,. \tag{37}$$

Given the likelihood function and the independent gamma priors on the defection and purchase rates, the joint posterior distribution of the behavioral parameters can be written as

$$\begin{aligned}\pi(\lambda,\mu|r,\alpha,s,\beta,x,t_x) &\propto f(x,t_x|\lambda,\mu)g(\lambda|r,\alpha)h(\mu|s,\beta)\\ &\propto \frac{\lambda^x}{\lambda+\mu}(\mu e^{-(\lambda+\mu)t_x}+\lambda e^{-(\lambda+\mu)T})\lambda^{(r-1)}e^{-\alpha\lambda}\mu^{(s-1)}e^{-\beta\mu}\,.\end{aligned} \tag{38}$$

Note that we consider the hyperparameters $(r,\alpha,s,\beta)$ to be fixed. The candidate draws in our random-walk Metropolis-Hastings sampler are generated using

$$\begin{aligned}\lambda^c &= \exp(\log\lambda+\varepsilon_\lambda), \qquad \varepsilon_\lambda \sim N(0,\sigma_\lambda^2)\\ \mu^c &= \exp(\log\mu+\varepsilon_\mu), \qquad \varepsilon_\mu \sim N(0,\sigma_\mu^2).\end{aligned}$$

In this way we ensure that the parameters always remain positive.

The parameters are now drawn sequentially using the following two-step Gibbs sampler:

1. Start sampling with initial values for $\lambda$ and $\mu$
2. Update $\lambda$
   - Draw the candidate value: $\lambda^c$
   - Compute $\alpha=\min\left(1,\pi(\lambda^c,\mu|r,\alpha,s,\beta,x,t_x)/\pi(\lambda,\mu|r,\alpha,s,\beta,x,t_x)\right)$.
   - With probability $\alpha$, set $\lambda=\lambda^c$
3. Update $\mu$:
   - Draw the candidate value: $\mu^c$
   - Compute $\alpha=\min\left(1,\pi(\lambda,\mu^c|r,\alpha,s,\beta,x,t_x)/\pi(\lambda,\mu|r,\alpha,s,\beta,x,t_x)\right)$.
   - With probability $\alpha$, set $\mu=\mu^c$
4. Repeat steps 2 and 3.

## Appendix B.2 BG/NBD model

For the conditional posterior distribution of the transaction rate, we have $\pi(\lambda|x,t_x,p) \propto \pi(\lambda,p)\pi(x,t_x|\lambda,p)$ such that

$$\pi(\lambda|x,t_x,p) \propto \lambda^{x+r-1} \times \begin{cases} p\,e^{-\lambda(t_x+\alpha)} + (1-p)e^{-\lambda(T+\alpha)} & \text{if } 0 < t_x \leq T \\ e^{-\lambda(T+\alpha)} & \text{if } x = 0 = t_x\,. \end{cases}$$

We, therefore, have

$$\pi(\lambda|x,t_x,p) = \frac{\frac{p}{(t_x+\alpha)^{x+r}}}{\frac{p}{(t_x+\alpha)^{x+r}} + \frac{1-p}{(T+\alpha)^{x+r}}} \varphi_{x+r,t_x+\alpha}(\lambda) + \frac{\frac{1-p}{(T+\alpha)^{x+r}}}{\frac{p}{(t_x+\alpha)^{x+r}} + \frac{1-p}{(T+\alpha)^{x+r}}} \varphi_{x+r,T+\alpha}(\lambda), \tag{39}$$

where $\varphi_{x,\beta}$ is the density of a gamma distribution with shape parameter $x$ and rate parameter $\beta$.

Likewise, for the conditional posterior distribution of the defection probability, we have

$$\pi(p|x,t_x,\lambda) \propto \pi(\lambda,p)\pi(x,t_x|\lambda,p) \propto \begin{cases} p^a(1-p)^{b+x-2}e^{-\lambda t_x} + p^{a-1}(1-p)^{b+x-1}e^{-\lambda T} & \text{if } 0 < t_x \leq T \\ p^{a-1}(1-p)^{b-1} & \text{if } x = 0 = t_x \end{cases}$$

and so

$$\pi(p|x,t_x,\lambda) = \frac{a}{a+(b+x-1)e^{-\lambda(T-t_x)}} \beta_{a+1,b+x-1}(p) + \frac{(b+x-1)e^{-\lambda(T-t_x)}}{a+(b+x-1)e^{-\lambda(T-t_x)}} \beta_{a,b+x}(p) \tag{40}$$

where $\beta_{a,b}$ is the density of a beta distribution with parameters $a$ and $b$.

## Appendix B.3 PDO model

For the conditional posterior distribution of the transaction rate in the PDO model, we get

$$\pi(\lambda|x,t_x,p) \propto \pi(\lambda,p)\pi(x,t_x|\lambda,p) \propto p \sum_{n=m_x}^{N} \frac{(1-p)^{n-1}}{(\alpha+(n-1)\tau)^x} \varphi_{x+r,\alpha+(n-1)\tau}(\lambda) + \frac{(1-p)^N}{(\alpha+T)^x} \varphi_{x+r,\alpha+T}(\lambda),$$

so that

$$\pi(\lambda|x,t_x,p) = \sum_{n=m_x}^{N} \frac{w_{x,p}^{(n)}}{W_{x,t_x,p}} \varphi_{x+r,\alpha+(n-1)\tau}(\lambda) + \frac{w_{x,p}^{(N+1)}}{W_{x,t_x,p}} \varphi_{x+r,\alpha+T}(\lambda), \tag{41}$$

where

$$w_{x,p}^{(n)} = \begin{cases} p\,\frac{(1-p)^{n-1}}{(\alpha+(n-1)\tau)^{x+r}} & \text{if } 1 \leq n \leq N \\ \frac{(1-p)^N}{(\alpha+T)^{x+r}} & \text{if } n = N+1, \end{cases}$$

and $W_{x,t_x,p} = \sum_{n=m_x}^{N+1} w_{x,p}^{(n)}$

For the conditional posterior distribution of the defection probability, it holds

$$\pi(p|x,t_x,\lambda) \propto \pi(\lambda,p|x,t_x) \propto \pi(\lambda,p)\pi(x,t_x|\lambda,p) \propto p^a \sum_{n=m_x}^{N} (1-p)^{b+n-2} e^{-\lambda(T-(n-1)\tau)} + p^{a-1}(1-p)^{b+N-1}\,.$$

Therefore,

$$\pi(p|x,t_x,\lambda) = \sum_{n=m_x}^{N} \frac{v_\lambda^{(n)}}{V_{t_x,\lambda}} \beta_{a+1,b+n-1}(p) + \frac{v_\lambda^{(N+1)}}{V_{t_x,\lambda}} \beta_{a,b+N}(p), \tag{42}$$

is a mixture of beta distributions where

$$v_\lambda^{(n)} = \begin{cases} B(a+1,b+n-1)e^{-\lambda(T-(n-1)\tau)} & \text{if } m_x \le n \le N \\ B(a,b+N) & \text{if } n = N+1\,. \end{cases}$$

and $V_{t_x,\lambda} = \sum_{n=m_x}^{N+1} v_\lambda^{(n)}$ and $B(\cdot,\cdot)$ is the beta function. Note that the value $V_{t_x,\lambda}$ depends on the data only through $m_x$.

# Appendix C HB estimation with a very diffuse prior on CDNOW dataset

Table 15 presents the mean of unconditional expectations for the CDNOW data under a very diffuse prior distribution. Recall that the prior parameters are chosen as $\nu_0 = J+3$ and $\Gamma_0 = \nu_0 I$, where $J$ represents the number of parameters of a customer (see Rossi, Allenby, and McCulloch (2005, Page 30)).

Table 15: Average of unconditional expectations in calibration period - under a diffuse prior on CDNOW data

| | HB1 | HB2 | HB3 | HB4 |
|---|---|---|---|---|
| Avg. $\mathbb{E}[x]$ | 0.228 | 0.096 | 0.253 | 0.209 |
| Avg. $\mathbb{E}[t_x]$ | 2.852 | 1.110 | 3.151 | 2.654 |

Although a very diffuse prior leads to badly estimated individual-level parameters, this does not necessary lead to bad predictions on the future transaction number and the timing predictions. The main reason for this is that these metrics are bounded. Figure 5 and Tables 16 to 19 show the forecasting performance of the HB models under this very diffuse prior. Hence, it is important to also look at the posterior distributions of the individual-level parameters. As noted earlier, these are very extreme under a diffuse prior for this data set.

Figure 5: Conditional expectation of future transaction frequency and future transaction timing on CDNOW - under a diffuse prior

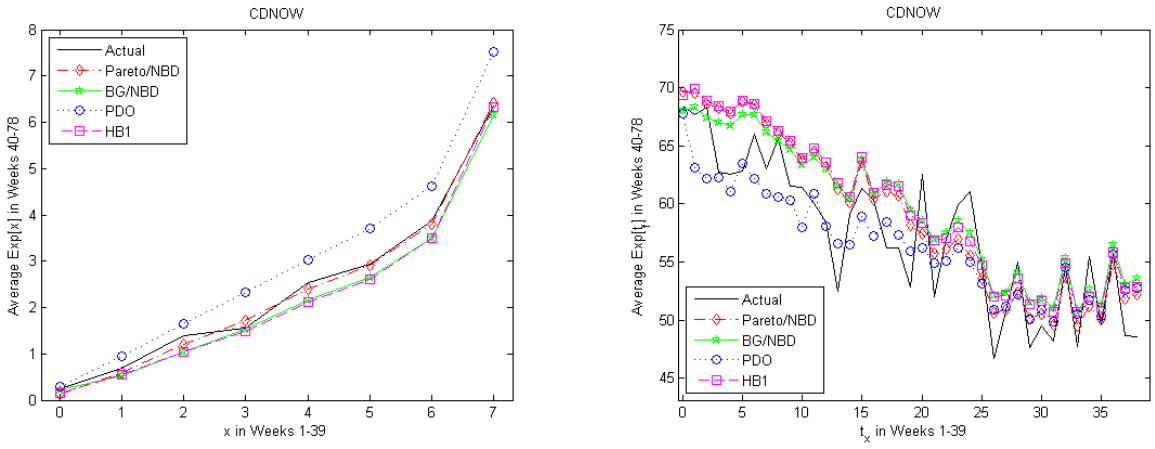



Table 16: In-sample predictive performance for unconditional predictions of the expected number of transactions and expected timing of last transaction - under a diffuse prior on CDNOW data

| | | $\mathbb{E}[x]$ | | | | $\mathbb{E}[t_x]$ **-weeks-** | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | MSE | MAE | ME+ | ME− | MSE | MAE | ME+ | ME− |
| CDNOW | HB1 | 5.454 | 1.087 | 0.228 | 2.369 | 130.586 | 7.537 | 2.747 | 16.236 |
| | HB2 | 5.689 | **1.061** | 0.096 | 2.501 | 147.785 | **7.081** | 1.094 | 16.653 |
| | HB3 | 5.414 | 1.092 | 0.253 | 2.344 | 128.279 | 7.626 | 3.024 | 16.172 |
| | HB4 | 5.486 | 1.083 | 0.208 | 2.388 | 132.239 | 7.481 | 2.556 | 16.357 |

Table 17: Model's prediction performance on the number of transactions - under a diffuse prior on CDNOW data

| | | Correlation | MSE | MAE | ME+ | ME− |
|---|---|---|---|---|---|---|
| CDNOW | HB1 | 0.6245 | 2.606 | 0.758 | 0.413 | 1.858 |
| | HB2 | 0.6154 | 2.890 | 0.748 | 0.302 | 1.990 |
| | HB3 | 0.6185 | 2.997 | 0.795 | 0.523 | 1.962 |
| | HB4 | 0.6173 | 2.744 | **0.680** | 0.247 | 2.094 |

Table 18: Highest Posterior Density Region and mean of correlations between behavioral rates - under a diffuse prior on CDNOW data

| | $\rho_{\theta_\lambda\theta_\mu}$ | | | $\rho_{\theta_\lambda\theta_\eta}$ | | | $\rho_{\theta_\eta\theta_\mu}$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | HPDR | | mean | HPDR | | mean | HPDR | | mean |
| CDNOW | -0.163 | 0.297 | 0.078 | 0.070 | 0.312 | 0.188* | -0.868 | -0.835 | -0.853* |

Table 19: Model's prediction performance on the time of next transaction - under a diffuse prior on CDNOW data

| | Correlation | MSE | MAE | ME+ | ME− |
|---|---|---|---|---|---|
| HB1 | 0.5770 | 126.257 | 7.502 | 17.232 | -4.052 |
| HB2 | 0.5538 | 291.028 | 16.054 | 9.423 | -17.628 |
| HB3 | 0.5491 | 142.779 | **6.494** | 5.329 | -10.314 |
| HB4 | 0.5665 | 271.112 | 15.367 | 9.053 | -16.873 |

# References


Abe, M. (2009a). "Counting Your Customers One by One: A Hierarchical Bayes Extension to the Pareto/NBD Model." In: *Marketing Science* 28.3, pp. 541–553.

– (2009b). "Customer Lifetime Value and RFM Data: Accounting Your Customers: One by One." In: *Working paper. Available from: <http://repository.dl.itc.u-tokyo.ac.jp/dspace/handle/2261/25752>. Accessed: 01.01.2010.*

Batislam, E.P., M. Denizel, and A. Filiztekin (2007). "Empirical validation and comparison of models for customer base analysis." In: *International Journal of Research in Marketing* 24.3, pp. 201–209.

Bueschken, J. and S. Ma (2012). "When are Your Customers Active and is Their Buying Regular or Random? An Erlang Mixture State-Switching Model for Customer Scoring." In: *Working paper, available at SSRN 2006410. Accessed: 01.08.2012.* URL: `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2006410`.

Ehrenberg, A.S.C. (1988). *Repeat-buying: Facts, theory and applications*. Griffin New London.

Fader, P.S., B.G.S. Hardie, and K.L. Lee (2005a). "Counting Your Customers" the Easy Way: An Alternative to the Pareto/NBD Model." In: *Marketing Science*, pp. 275–284.

– (2005b). "RFM and CLV: Using iso-value curves for customer base analysis." In: *Journal of Marketing Research*, pp. 415–430.

Geweke, J. et al. (1991). *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*. Federal Reserve Bank of Minneapolis, Research Department.

Gilks, W.R., S. Richardson, and D.J. Spiegelhalter (1996). *Markov chain Monte Carlo in practice*. Chapman & Hall/CRC.

Gönül, F.F. and F.T. Hofstede (2006). "How to compute optimal catalog mailing decisions." In: *Marketing Science*, pp. 65–74.

Gupta, S. (1988). "Impact of sales promotions on when, what, and how much to buy." In: *Journal of Marketing Research*, pp. 342–355.

Gupta, S. et al. (2006). "Modeling customer lifetime value." In: *Journal of Service Research* 9.2, pp. 139–155.

Hastings, W.K. (1970). "Monte Carlo Sampling Methods Using Markov Chains and Their Applications." In: *Biometrika*, pp. 97–109.

Hyndman, Rob J. (1996). "Computing and Graphing Highest Density Regions." In: *The American Statistician* 50.2, pp. 120–126.

Jerath, K., P.S. Fader, and B.G.S. Hardie (2011). "New Perspectives on Customer 'Death' Using a Generalization of the Pareto/NBD Model." In: *Marketing Science* 30.5, pp. 866–880.

Morwitz, V.G. and D. Schmittlein (1992). "Using segmentation to improve sales forecasts based on purchase intent: Which" intenders" actually buy?" In: *Journal of Marketing Research*, pp. 391–405.

Reinartz, W.J. and V. Kumar (2000). "On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing." In: *The Journal of Marketing*, pp. 17–35.

– (2003). "The impact of customer relationship characteristics on profitable lifetime duration." In: *Journal of Marketing*, pp. 77–99.

Rossi, P., G.M. Allenby, and R. McCulloch (2005). *Bayesian statistics and marketing*. John Wiley and Sons, Ltd.

Schmittlein, D.C., D.G. Morrison, and R. Colombo (1987). "Counting your customers: Who are they and what will they do next?" In: *Management Science*, pp. 1–24.

Schmittlein, D.C. and R.A. Peterson (1994). "Customer base analysis: An industrial purchase process application." In: *Marketing Science*, pp. 41–67.

Schmittlein, David C., Lee G. Cooper, and Donald G. Morrison (1993). "Truth in Concentration in the Land of (80/20) Laws." In: *Marketing Science* 12.2, pp. 167–183.

Schwartz, E., E. Bradlow, and P. Fader (2012). "Model Selection Using Database Characteristics: Classification Methods and an Application to the'HMM and Its Children'." In: *Working paper, available at SSRN 2085767. Accessed: 01.08.2012.* URL: `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2085767`.

Talluri, K.T. and G. Van Ryzin (2005). *The theory and practice of revenue management*. Vol. 68. Springer Verlag.

Tereyağoğlu, N., P. Fader, and S. Veeraraghavan (2012). "Filling Seats at a Theater: Estimating the Impact of Posted Prices and Dynamic Discounts." In: *Working paper. Accessed: 01.10.2012.* URL: `http://opim.wharton.upenn.edu/~senthilv/papers/Filling_seats.pdf`.

Wu, C. and H.L. Chen (2000). "Counting your customers: Compounding customer's in-store decisions, interpurchase time and repurchasing behavior." In: *European Journal of Operational Research* 127.1, pp. 109–119.

Wübben, M. and F. Wangenheim (2008). "Instant customer base analysis: Managerial heuristics often 'get it right'." In: *Journal of Marketing* 72.3, pp. 82–93.

Zhang, J. and L. Krishnamurthi (2004). “Customizing promotions in online stores.” In: *Marketing Science*, pp. 561–578.