GTBMC 2022
Global Technology and Business Management Conference

A Key Model Enhancement for Personalized Marketing

Introducing Purchase Timing Predictions to Hierarchical Bayes Pareto/NBD Model

1st Evşen Korkmaz
Technology & Operations Management,
Rotterdam School of Management
Erasmus University Rotterdam
The Netherlands
evsenko@gmail.com or 0000-0002-2886-0462

2nd Dennis Fok
Econometrics Department,
Erasmus School of Economics
Erasmus University Rotterdam
The Netherlands
dfok@ese.eur.nl

3rd Roelof Kuik
Technology & Operations Management,
Rotterdam School of Management
Erasmus University Rotterdam
The Netherlands
rkuik@rsm.nl

***Abstract*— Personalized marketing is a key business strategy that uses customer data to offer an optimized marketing experience. Leveraging the extensive customer data, businesses can identify behavioral patterns to more effectively design targeted marketing tactics. Buy-till-you-defect [BTYD] models are one of the key enablers in this endeavor as they specify customers' transaction and defection processes for businesses operating under a non-contractual setting. These models have been typically used to identify active customers in a company's customer-base as well as to predict purchase frequency and amount. Given the rise of personalized marketing, companies continuously need to improve their understanding of customers to stay ahead of competition. In this article, we enhanced BTYD models' predictive capability so that these models can jointly predict also the customer's decision of *when* to shop, together with how often to shop and how much to spend. Purchase timing predictions are managerially relevant as they enable marketing executives to choose appropriate targeting and promotion strategies to improve revenues. For two well established BTYD models, Pareto/NBD model and its Hierarchical Bayes extension, we derive closed-form expressions for future purchase timing. Next, we validate these timing predictions on real datasets. We believe extending the use of BTYD models with this additional model output will lead to higher business adoption.**

Keywords—Buy–till–you–defect models; purchase timing; Bayesian estimation; customer base analysis; personalization

I. Introduction

Many firms routinely store extensive customer transaction data. However, processing this data in order to design personalized customer journeys can still be a challenge. The customer base analysis literature provides a number of methods to leverage such rich data to predict customer's transaction behavior. In literature, a distinction is made between a contractual and a non-contractual setting. The latter is especially challenging as one does not observe the moment at which an individual stops being a customer of 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 quit buying from the company. A wide variety of models is available for these purposes.

The ever-growing online retail industry is an important example of the non-contractual setting. Retailers never know which customers are still active, or in other words, will continue buying from the company. Thus, the customer database of an online retailer is likely to contain many inactive customers. For example, in 2005 eBAY reported 168 million registered customers but only 68 million of them were counted as active by the company [1]. It is, therefore, very useful to develop a method to identify active customers under a non-contractual setting.

It has been widely recognized in the literature that models that ignore defection, like the early NBD model by A.S.C. Ehrenberg [2], do not provide good predictions for this type of industry. They generally overestimate future transaction frequencies as demonstrated in [3]. The first model that does account for defection was proposed by D.C. Schmittlein and R.A. Peterson [4]. Since then, there has been a strong focus on the so-called buy-till-you-defect (BTYD) model. Several extensions of this model have been introduced, see [5], [6] and [7]. Some of these models have also been used to generate managerially relevant insights as in [8], [9] and [10].

In this paper, we discuss that another key metric on customer behavior can also be jointly predicted by these models. This metric is the highly relevant future purchase timing. These timing predictions critically depend on the interplay between the assumed transaction and defection processes. Yet, predicting the timing of the next purchase is not straightforward. We develop methods for the state-of-the-art BTYD models to enable them to also deliver such timing predictions.

We present out-of-sample performances in predicting the transaction timing of each customer for three datasets. The first dataset is from an online grocer in a Western European country. The second is the well-known CDNOW dataset which has been commonly used as a benchmark set. The third dataset is also used by K. Jerath, P.S. Fader and B.G.S. Hardie [7] and by E.P. Batislam, M. Denizel, and A. Filiztekin [11], and is from a Turkish grocery retailer.

Our results show that BTYD models can predict purchase timing to a great extent. We also discuss that certain data characteristics a-priori indicate how well the model predict.

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

II. Buy-till-you-defect Models

In this section, we briefly introduce the main ideas underlying the BTYD models. After discussing the similarities and differences across the models in scope, we articulate our key contribution.

A. BTYD Modeling Literature

The Pareto/Negative Binomial Distribution (Pareto/NBD) model is the first model that considers the customer's defection process [4]. This model assumes that, while active, customers make purchases according to a Poisson process with heterogeneous rates. Customer defection 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 so far has enabled 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 literature. As suggested by P.S. Fader, B.G.S. Hardie, and K.L. Lee [5], the continuous time defection process is replaced by a discrete time process. More specifically, after each purchase the customer defects with an individual-specific probability. The resulting model is called 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, the Periodic-Death-Opportunity (PDO) model was introduced by K. Jerath, P.S. Fader, and B.G.S. Hardie [7]. 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 Pareto/NBD model, and also in the above-mentioned extensions, these rates are assumed to be independent. In practice, this assumption may be too restrictive hence violated as frequent shoppers tend to have a long lifetime. This would imply a negative correlation between both rates. A Hierarchical Bayes extension of the Pareto/NBD model that incorporates such correlation is suggested in by M. Abe [6]. In this model, gamma heterogeneity 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 leveraged to gain further insights on customer behavior. A disadvantage of this extension is that for some quantities, closed-form expressions are no longer available. As a result, the proposed model in [6] needs Bayesian (simulation) techniques. We will refer to this model as the HB model.

In this paper, we focus on the high-performing Pareto/NBD model and its Hierarchical Bayes (HB) extension.

B. Our contribution

We show that another key behavioral dimension can also be jointly predicted using the same models. We derive specific formulas and present a general method to predict the timing of the next purchase for the two established BTYD models. Given the memoryless property on inter-arrival 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.

We also validate the newly introduced individual level timing predictions on different datasets and discuss their managerial implications.

III. Introducing the Timing of Transactions

In this section, we present the models in technical terms. Both models in scope 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 θ_i . On the population-level, all models specify a heterogeneity distribution for (the elements of) θ_i. This distribution is parameterized by hyper-parameters which are denoted by ξ.

Table I gives a summary of the assumptions and the dominant estimation method for both models. We distinguish between assumptions on individual behavior and on heterogeneity modeling. Both models in scope have the same assumption on the purchase process of an individual, while active as well as on defection process; they differ in the way how they model heterogeneity.

TABLE I. MODEL COMPARISON WITH RESPECT TO THE ASSUMPTIONS AND ESTIMATION PROCESS

	Pareto/NBD	*Hierarchical Bayes*
Purchase Process Defection Process Defection Timing	Poisson Exponential Continuous	Poisson Exponential Continuous
Purchase rate distribution Defection rate distribution	Gamma Gamma	Bi-variate log-normal
Estimated parameters Estimation procedure	Hyper params. MLE	Hyper & indiv. params MCMC

In the remainder of this section, we first present closed-form expressions for the last transaction timing in the calibration period and the first transaction timing in the holdout period and then discuss the sampling of (the hyperparameters and) the behavioral parameters for both models. Note that, we are the first to derive these expressions.

A. 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 μ_i. While active, this customer makes purchases according to a Poisson process with rate λ_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 λ_i has shape parameters r, and scale parameter α. The shape and scale parameters for μ_i are s and β, 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.

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) are presented in [4].

We demonstrate that the Pareto/NBD model can be leveraged 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 λ_i and μ_i, we can obtain the expected timing of the last purchase as

$$E\left(t_{x,i} \middle| \lambda_i, \mu_i, T_i\right) = \frac{1-e^{-\mu_i T_i}}{\mu_i} - \frac{1-e^{-(\lambda_i+\mu_i)T_i}}{\lambda_i+\mu_i} \tag{1}$$

see Appendix A for the associated derivations. Note that we are the first to make these derivations on purchase timing. By comparing $E[t_{x,i} \mid \lambda_i, \mu_i, T_i]$, averaged over the estimated distribution of λ_i and μ_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 already defected. 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, we show that the conditional expectation of $t_{f,i}$ in the Pareto/NBD model equals

$$E\left(t_{f,i} \middle| x_i, t_{x,i}, T_i, \lambda_i, \mu_i\right) = \left(1 - P[t_{\Delta,i} > T_i | x_i, t_{x,i}, T_i, \lambda_i, \mu_i]\right)T^+ + P[t_{\Delta,i} > T_i | x_i, t_{x,i}, T_i, \lambda_i, \mu_i]\left(T_i + \frac{1-e^{-(\lambda_i+\mu_i)(T_i^+-T_i)}}{\lambda_i+\mu_i}\right) \tag{2}$$

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

$$\frac{\lambda_i}{\lambda_i + \mu_i e^{(\lambda_i+\mu_i)(T_i - t_{x,i})}} \tag{3}$$

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, $\xi=(\alpha, r, \beta, s)$:

$$\begin{aligned}\pi(\theta_i | data_i, \xi) &= \pi(\lambda_i, \mu_i | r, \alpha, s, \beta, \Gamma, x_i, t_{x,i}, T_i) \\ &\propto f(x_i, t_{x,i} | \lambda_i, \mu_i) g(\lambda_i | r, \alpha) h(\mu_i | s, \beta) \\ &\propto \frac{\lambda_i^{x_i}}{\lambda_i + \mu_i} \left(\mu_i e^{-(\lambda_i+\mu_i)t_{x,i}} + \lambda_i e^{-(\lambda_i+\mu_i)T_i}\right) \frac{\alpha^r}{\Gamma(r)} \lambda_i^{(r-1)} e^{-\alpha\lambda_i} \frac{\beta^s}{\Gamma(s)} \mu_i^{(s-1)} e^{-\beta\mu_i}\end{aligned} \tag{4}$$

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, $\pi(\theta_i | data_i, \xi)$ where, again, θ_i denotes the individual-level parameters for individual *i* and ξ denotes the hyperparameters associated with the whole customer base in the BTYD model. A Metropolis-Hastings algorithm can be used to sample from this posterior density. Details of this sampling algorithm are presented in Appendix B.

B. *Hierarchical Bayes Extension of the Pareto/NBD Model*

The Pareto/NBD model presented above does not allow the individual-level parameters to be correlated and it does not take into account customer characteristics. In many cases, individual-level characteristics are available and may be useful in predicting customer behavior. Therefore, a Hierarchical Bayes [HB] extension of the Pareto/NBD model is proposed in by M. Abe in [6] where 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. The use of Markov chain Monte Carlo [MCMC] techniques to estimate the (hyper)parameters and to calculate various metrics are proposed in [6].

Same individual-level assumptions are made in [6] 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 $N(\beta,\Gamma)$. Γ 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

$$\pi(\{x_i, t_{x,i}, T_i, \lambda_i, \mu_i\}, \beta, \Gamma) = \prod_{i=1}^{N}\left(\pi(x_i, t_{x,i} | \lambda_i, \mu_i)\pi(\lambda_i, \mu_i | \beta, \Gamma)\right) \pi(\beta, \Gamma). \quad (4)$$

Here $\pi(\beta,\Gamma)$ is the prior distribution of the population-level parameters β and Γ. The standard conjugate prior is used, that is, $\beta \sim N(\beta_0, A_0)$ and Γ follows an inverted Wishart distribution with parameters (v_0, Γ_0). As the individual-level behavioral assumptions of the HB model are identical to the Pareto/NBD model, conditional on λ_i and μ_i, all timing related expressions are the same. Draws for the individual-level parameters are a natural by-product of the MCMC sampler.

An extension of the HB model is proposed in [12] by adding a customer behavioral characteristic which is the amount of spending. Hereby, the individual parameter vector, θ_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 study. Consequently, we consider four different configurations of the HB model. The first configuration (HB1) represents the HB model without any covariates. 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 [13]. 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 [6]. As our sampler differs from the one presented in [6], we present the main steps of the sampler below:

[0] Set initial value for θ_i, $i = 1,...,N$.
[1a] Generate $z_i \mid t_{x,i}, x_i, T_i, \theta_i$ according to the being active probability given in Equation (3), for $i = 1,...,N$.
[1b] If $z_i = 0$, generate $t_{\delta,i} \mid t_{x,i}, x_i, T_i, z_i, \theta_i$ using an exponential distribution truncated to $(t_{x,i}, T_i)$.
[2] Generate $\beta,\Gamma \mid \theta_i$ *for* $i=1,\dots N$ using a standard multi-variate normal regression update (see [14] page 34).
[3] Generate $\theta_i \mid t_{x,i}, x_i, T_i, z_i, t_{\Delta,i}, \beta, \Gamma$ with a Gaussian random-walk MH algorithm, for $i = 1,...,N$.

The step size in the random-walk MH algorithm is set by applying an adaptive MH method in the burn-in phase, see further details in [15].

IV. DATA

We leverage three real datasets to predict and validate the purchase timing. The first dataset contains daily transaction data of an online grocery retailer in a Western European country. 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 transaction data of 2357 customers who made their first purchase 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 dataset is also used in [7] and [11]. 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 studies, we use the first 78 weeks for calibration and leave 13 weeks for validation purposes.

Despite the similar format of transactional data across these three datasets, there are significant differences. First and foremost, the companies that supply the data operate in different industries, i.e. grocers vs entertainment. While two of them are online retailers, one is a brick-and-mortar retailer with a strong loyalty program. We believe such differences are important to be considered while building managerial insights.

V. Empirical Results

In this section, we focus on individual-level predictions conditional on the individual's history and present results in predicting the timing of the first out-of-sample purchase.

Note that for the online retailer datasets (online grocer and CDNOW), covariate data on the average number of shopping items per customer is available. Hence 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, θ_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 applied. 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, see how to evaluate the accuracy of sampling-based approaches to the calculation of posterior metrics in [17].

As discussed in Section III, for some metrics of interest, obtaining closed-form expression conditioned on an individual's history and hyperparameters can be extremely cumbersome. We, therefore, first obtain draws for the individual's behavioral parameters from the posterior densities and next calculate the expected value of 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 iterations 300,000 times, of which only the last 10,000 iterations were used.

We focus on the performance on predicting future transaction timing.[1] 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.

Table II presents an overview of the main results. The HB models perform rather well on the grocer and online grocer datasets. This can be explained by the fact that we found a significant correlation between the behavioral parameters for both datasets. As explained earlier, Pareto/NBD model assumes independence between purchase and defection rates, whereas HB models relax this assumption by allowing these individual-level parameters to follow a bivariate log-normal distribution.

Pareto/NBD model outperforms the HB models on CDNOW dataset in out-of-sample predictions. CDNOW data exhibits certain characteristics that are not valid for the two other datasets, such as relatively low number of repeat transactions as well as no dependency between behavioral parameters.

TABLE II. Models' prediction performance on the timing of next transaction

Dataset	*Model*	*Correlation*	*MSE*	*MAE*
Online Grocer	Pareto/NBD	0.7296	46.674	4.508
	HB1	**0.7328**	**43.416**	4.223
	HB2	0.7254	44.374	4.296
	HB3	0.7201	46.594	**4.067**
	HB4	0.7204	46.504	4.073
CDNOW	Pareto/NBD	**0.5789**	**125.451**	**7.372**
	HB1	0.5486	273.555	15.660
	HB2	0.5449	282.423	15.865
	HB3	0.5687	270.514	15.408
	HB4	0.5689	270.028	15.376
Grocer	Pareto/NBD	0.8183	7.684	1.442
	HB1	**0.8190**	**7.602**	**1.426**

Among the HB models, a remarkable point is the improved performance of the HB3/4 models that consider the average spending amount on CDNOW and online grocer datasets. We can explain this with the existence of the strong and significant negative correlation between the spending amount and defection parameters in both datasets.

Note that we observe an interesting point solely on the online grocer data that MSE and MAE measures favor different versions of HB models. This can be explained by the fact that there are a limited number of outliers with the HB3/4 models. This results in having a higher MSE for HB3 model as this metric is more sensitive to outliers, i.e. the error increases a quadratic fashion in MSE versus in a proportional fashion in MAE.

VI. Discussion

In this paper, our goal is to enhance the Pareto/NBD model and its HB extension with additional predictive capabilities. We argue that prediction of the future transaction timing of an individual is a managerially relevant metric and show that leveraging the exact same model we can jointly predict the timing of next transactions together with future transaction frequency and amount.

Timing predictions help to optimize several marketing strategies, such as setting price, promotion and advertising policies as discussed by K. Jedidi, C. F. Mela, and S. Gupta in [18]. For example, consider an online retailer implementing micro-marketing strategies. The most appropriate time to contact its customers depends on their

[1] We thank E.P. Batislam, M. Denizel, and A. Filiztekin [11] and P.S. Fader, B.G.S. Hardie, and K.L. Lee [16] for making the out-of-sample timing data available.

expected timing of the next purchase. High quality timing predictions may contribute to achieving the full potential of micro-marketing [19].

Following the pioneering research by S. Gupta [20], there is a growing literature that examines the effectiveness of promotions on whether to buy, *'when'* to buy, and how much to buy; see a summary of relevant literature in [21]. We believe that using these 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, with individual level timing predictions, an online retailer can optimize its delivery fleet capacity or even optimize its delivery schedules on a daily basis. Given the fact that online retailers have limited delivery capacity at a given time, operations managers can also prioritize key customer groups for highly demanded delivery time slots based on these predictions [22]. Overall, as emphasized by N. Tereyagoglu, P. Fader, and S. Veeraraghavan, having accurate timing predictions has a crucial role in improving revenues [23].

Besides the relevance of timing predictions from operations and marketing perspectives, let us now discuss why it is important for businesses to get more predictive output from a single model. Despite the advances in enterprise level machine learning model automation, businesses still trust one integrated model that can jointly predict multiple metrics such as purchase frequency, amount and timing, over multiple models that predict single metrics independently. Hence it is commonly observed that the more behavioral dimensions a model can predict, the more likely it is to be adopted by business. As discussed by P. Boatwright, S. Borle and J. B. Kadane, the interest in models that can predict key quantities in a joint, dependent manner also depends on which type of industry the business operates in [24]. The authors argue that especially online retailers need a joint model for different dimensions of customers' purchase behavior, such as timing and frequency.

In summary, we believe that the ability to predict the timing of future transactions can be instrumental in accelerating not only BTYD models' adoption by business, but also research on predictive performance of these models across different industries such as apparel, home appliances and electronics retailing.

Overall, we present a general method and derive specific formulas that can be used to predict the timing of the next purchases for two established BTYD models. We are the first to derive these formulas. We use these methods to compare the predictive performance of the models in scope on three very different datasets. We advocate further testing of these predictions on other datasets to expand our understanding of these models' capability and to build industry specific insights.

References

[1] S. Gupta, D. Hanssens, B. Hardie, W. Kahn, V. Kumar, N. Lin, N.R.S. Sriram, "Modeling customer lifetime value," *J. Serv. Res*, vol. 9, no. 2, pp. 139– 155, Nov. 2006.

[2] A.S.C. Ehrenberg, *Repeat-buying: Facts, theory and applications*, London: Oxford University Press, 1988.

[3] D. C. Schmittlein and R. A. Peterson, "Customer base analysis: An industrial purchase process application," *Mark. Sci.*,vol. 13, no. 1, pp. 41–67, 1994.

[4] Schmittlein, D.C., D.G. Morrison, and R. Colombo, "Counting your customers: Who are they and what will they do next?," *Manag. Sci.*, vol. 33, no. 1, pp. 1–24, Jan. 1987.

[5] P. S. Fader, B. G. S. Hardie, and K. L. Lee, "Counting Your Customers" the Easy Way: An Alternative to the Pareto/NBD Model," *Mark. Sci.*, vol. 24, no. 2, pp. 275–284, 2005a.

[6] M. Abe, "Counting Your Customers One by One: A Hierarchical Bayes Extension to the Pareto/NBD Model," *Mark. Sci.*, vol. 28, no. 3, pp. 541–553, 2009a.

[7] K. Jerath, , P. S. Fader, and B. G. S. Hardie, "New Perspectives on Customer 'Death' Using a Generalization of the Pareto/NBD Model," *Mark. Sci.*, vol. 30, no. 5, pp. 866–880, 2011.

[8] W. J. Reinartz and V. Kumar, "On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing," *J. Mark.*, vol. 64, no. 4, pp. 17–35, Oct. 2000.

[9] W. J. Reinartz and V. Kumar, "The impact of customer relationship characteristics on profitable lifetime duration," *J. Mark.*, vol. 67, no. 1, pp. 77–99, Jan. 2003.

[10] M. Wübben and F. Wangenheim, "Instant customer base analysis: Managerial heuristics often 'get it right'," *J. Mark.*, vol. 72, no. 3, pp. 82–93, May 2008.

[11] E. P. Batislam, M. Denizel, and A. Filiztekin, "Empirical validation and comparison of models for customer base analysis," *Int. J. Res. Mark.*, vol. 24, no. 3, pp. 201–209, Sep. 2007.

[12] M. Abe, "Customer Lifetime Value and RFM Data: Accounting Your Customers: One by One," Working paper, 2009b [Online] http://repository.dl.itc.u-tokyo.ac.jp/dspace/handle/2261/25752

[13] W. K. Hastings, "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, Apr. 1970.

[14] P. Rossi, G. M. Allenby, and R. McCulloch, *Bayesian statistics and marketing*, Hoboken: John Wiley and Sons, Ltd., 2005.

[15] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov chain Monte Carlo in practice*, Boca Raton: Chapman & Hall, 1996.

[16] P. S. Fader, B. G. S. Hardie, and K. L. Lee, "RFM and CLV: Using iso-value curves for customer base analysis," *J. Mark. Res.*, vol. 42, no. 4, pp. 415–430, Nov. 2005b.

[17] J. F. Geweke, "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments," Federal Reserve Bank of Minneapolis, Research Department, 1991.

[18] K. Jedidi, C. F. Mela, and S. Gupta, "Managing Advertising and Promotion for Long-Run Profitability," *Mark. Sci.*, vol. 18, pp. 1–22, 1999.

[19] J. Zhang and L. Krishnamurthi, "Customizing promotions in online stores," *Mark. Sci.*, pp. 561–578, Nov. 2004.

[20] S. Gupta, "Impact of sales promotions on when, what, and how much to buy," *J. Mark. Res.*, vol. 25, no. 4, pp. 342–355, Nov. 1988.

[21] F. F. Gönül and F. T. Hofstede, "How to compute optimal catalog mailing decisions," *Mark. Sci.*, vol. 25, no. 1, pp. 65–74, Jan. 2006.

[22] K. T. Talluri and G. Van Ryzin, *The theory and practice of revenue management*, New York, NY: Springer, 2005.

[23] N. Tereyagoglu, P. Fader, and S. Veeraraghavan, "Filling Seats at a Theater: Estimating the Impact of Posted Prices and Dynamic Discounts," Working paper, 2012, [Online] http://opim.wharton.upenn.edu/ ~senthilv/papers/Filling_seats.pdf

[24] P. Boatwright, S. Borle, and J. B. Kadane, "A Model of the Joint Distribution of Purchase Quantity and Timing," *J. Am. Stat. Assoc.*, vol. 98, no. 463, pp. 564-572, Sep. 2003.

APPENDICES

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

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.

Throughout this appendix, we drop the i subscript representing customer i for notational simplicity. For the sake of simplicity, we also do not condition on the length of the observational interval T.

The timing expressions are the same for both models as they have the same assumptions on individual behavior. The time of defection, t_Δ, has the probability function

$$\mathrm{P}(dt_\Delta|\lambda,\mu) = \mu e^{-\mu t_\Delta}\, dt_\Delta$$

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

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

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

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

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_δ. Integrating over t_δ, one obtains

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

Based on the equation above, the expected value on the time of the last transaction is calculated as follows,

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

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 : $E(t_f\,|x,t_x,\lambda,\mu)$. 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

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

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

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

Considering an active customer, the density of the first timing, t, of a transaction on (T,∞) is $\lambda e^{-(\lambda+\mu)(t-T)}$ and t has a point mass at infinity of $\mu/(\lambda+\mu)$ 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,

$$\mathrm{E}(t_f|x,t_x,\lambda,\mu) = p^+ \int_T^{T^+} t\,\pi_f\,(t\,|\,x,t_x, z=1,\lambda,\mu)dt + p^+\left(1 - \int_T^{T^+} \pi_f\,(t\,|\,x,t_x,z=1,\lambda,\mu)dt\right)T^+ + (1-p^+)T^+$$

$$\mathrm{E}(t_f|x,t_x,\lambda,\mu) = \mathrm{T} + \frac{\mu e^{(\lambda+\mu)(\mathrm{T}-t_x)} + \lambda e^{-(\lambda+\mu)\mathrm{T}}}{\lambda + \mu e^{(\lambda+\mu)(\mathrm{T}-t_x)}}(T^+ - T) + \frac{\lambda}{\lambda+\mu e^{(\lambda+\mu)(\mathrm{T}-t_x)}} + \frac{1-e^{-(\lambda+\mu)(T^+-T)}}{\lambda+\mu}$$

B. Estimation procedure for Pareto/NBD Model

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 model.

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.

The likelihood function for the Pareto/NBD model is

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

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

$$\pi(\lambda,\mu|r,\alpha,s,\beta,x,t_x) \propto f(x,t_x|\lambda,\mu)g(\lambda|r,\alpha)h(\beta|s,\beta) \propto \frac{\lambda^x}{\lambda+\mu}\left(\mu e^{-(\lambda+\mu)t_x} + \lambda e^{-(\lambda+\mu)T}\right)\lambda^{(r-1)}e^{-\alpha\lambda}\mu^{(s-1)}e^{-\beta\mu}$$

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

$$\lambda^c = \exp(\log\lambda + \varepsilon_\lambda), \quad \varepsilon_\lambda \sim N(0,\sigma_\lambda^2)$$

$$\mu^c = \exp(\log\mu + \varepsilon_\mu), \quad \varepsilon_\mu \sim N(0,\sigma_\mu^2)$$

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 λ and μ
2. Update λ
   - Draw the candidate value: λ^c
   - Compute α = min(1, $\pi(\lambda^c,\mu|r,\alpha,s,\beta,x,t_x)/\pi(\lambda,\mu|r,\alpha,s,\beta,x,t_x)$)
   - With probability α, set $\lambda = \lambda^c$
3. Update μ
   - Draw the candidate value: μ^c
   - Compute α = min(1, $\pi(\mu^c,\lambda|r,\alpha,s,\beta,x,t_x)/\pi(\lambda,\mu|r,\alpha,s,\beta,x,t_x)$)
   - With probability α, set $\mu = \mu^c$

Repeat steps 2 and 3.