International Journal of Research in Marketing 38 (2021) 194–215

Contents lists available at ScienceDirect

IJRM

International Journal of Research in Marketing

journal homepage: www.elsevier.com/locate/ijresmar

Full Length Article

Leveraging purchase regularity for predicting customer behavior the easy way

Thomas Reutterer[a,*], Michael Platzer[b], Nadine Schröder[a]

[a] Department of Marketing, WU Vienna University of Economics and Business, Welthandelsplatz 1, A-1020 Vienna, Austria
[b] MOSTLY AI Solutions, Hegelgasse 21/3, A-1010 Vienna, Austria

ARTICLE INFO

Article history:
First received on August 15, 2016 and was under review for 9½ months
Available online 12 September 2020

Area Editor: Bruce G.S. Hardie

Keywords:
Customer-base analysis
Probability models
Pareto/NBD
Pareto/GGG
Timing models
Regularity

ABSTRACT

The valuation of future customer activity is a mainstay of any organization seeking to efficiently manage its customer portfolio. In the area of customer-base analytics, the ongoing race for predictive power has yielded a large corpus of research to assist managers in this respect. Approaches in the tradition of stochastic models have been particularly successful because they rely only on easy-to-compute key metrics and integrate them within a parsimonious probability-modeling framework. Recent advances in this field have demonstrated that incorporating the timing regularity of past purchases can improve predictive accuracy relative to purely recency/frequency-based approaches. This paper expands that idea and introduces generalizations of a well-established probability model, the BG/NBD (Fader et al., 2005a), by replacing the exponential with a more flexible Erlang-k interarrival timing process. The resulting model variants are capable of leveraging regularity while retaining almost the same level of data requirements and algorithmic efficiency. Using extensive simulation studies and six data sets covering a wide range of empirical settings the authors demonstrate substantial improvements in predictive accuracy against the baseline models and performance gains close to or on par with a more complex model alternative. The availability of efficient and easily accessible implementations of the new model variants in the R-package BTYDplus allows marketing analysts to apply them in large-scale scenarios of data-rich environments on a continuous basis.

1. Introduction

The valuation of customers is at the core of any business seeking to optimize profitability by acquiring and retaining the most valuable customers. However, accurately assessing a customer's residual value (also referred to as "customer lifetime value", CLV) is based on her future activities and thus is unknown and needs to be estimated (Gupta et al., 2006). This becomes particularly challenging in settings in which the customer-firm relationship is of a noncontractual nature (Reinartz & Kumar, 2000), where the organization is unable to directly observe whether a customer is still active or has already (silently) defected. In any case, making predictions routinely (and subsequently assessing the predictive accuracy) for a large number of customers requires a systematic data-driven methodology. A variety of approaches developed in or around the marketing discipline ranges from regression-type models, via machine learning techniques to probabilistic models of consumer behavior (e.g., Fader & Hardie, 2009; Gupta

* Corresponding author.
E-mail addresses: thomas.reutterer@wu.ac.at, (T. Reutterer), michael.platzer@mostly.ai, (M. Platzer), nadine.schroeder@wu.ac.at. (N. Schröder).

https://doi.org/10.1016/j.ijresmar.2020.09.002
0167-8116/

et al., 2006; Malthouse, 2009). Nevertheless, any further improvements in the field of customer analytics will help marketers to exploit even better the heterogeneity in their customer base through tailored direct marketing activities.

Probability models are a particularly prominent model class for customer-base analysis in noncontractual settings. Most of these models rely on only two pieces of information, which can be easily computed from the transaction records of virtually any customer cohort, namely when a customer was last active (recency, R) and how active a customer has been in the past (frequency, F). A number of factors contribute to the ongoing popularity of RF-based probability models: They parsimoniously and efficiently utilize readily available data, they are rooted in well-grounded "behavioral stories", key variables of managerial interest are easy to estimate (because closed-form expressions are typically available), and – most important – they exhibit surprisingly good predictive accuracy for long decision horizons, at both the aggregate and the individual level (Batislam et al., 2007; Fader et al., 2005b; Fader et al., 2010; Fader, Hardie, Lee, 2005a). In an era of "big data," these arguments make RF-based models scalable for data-rich environments and thus a mainstay of any customer-base analyst's toolbox.

Despite their stunningly good performance in many empirical settings, some authors have recently challenged the capability of approaches using just two simple metrics to adequately summarize key aspects of a customer's purchase history. More specifically, Zhang et al. (2015) and Platzer and Reutterer (2016) make the case for additionally leveraging information on interevent timing patterns to gain predictive accuracy. Zhang et al. (2015) provide an easy-to-compute measure summarizing (ir-)regularities in timing patterns. The authors demonstrate that adding a clumpiness metric, which measures the degree of unevenly distributed interevent times, can improve RF-based regressions on out-of-sample observations. Platzer and Reutterer (2016) develop a fully integrated stochastic consumer behavior model that is capable of capturing a variety of timing patterns while still remaining robust in the presence of churn and for the, typically large, share of customers for whom few (fewer than ten) events have been recorded (see Section 5 of Platzer & Reutterer, 2016). This model, termed Pareto/GGG (i.e., a gamma-gamma-gamma mixture for modeling the interevent timing process and customer heterogeneity), results in improved predictive accuracy even for only mild degrees of regularity in the observed timing patterns. However, these gains in flexibility and predictive power come at the expense of a significant increase in computational cost and implementation complexity. For example, drawing individual-level Pareto/GGG parameters for a single cohort of 1.5 million customers newly acquired by an anonymous online retailer over a monthly time period requires 10 h of computation time.[1] Given anticipated advances in hardware technology and distributed computing power, this might not sound like a substantial issue at first sight. However, there are scenarios in which a fast, computationally efficient evaluation of future customer behavior is imperative. This is particularly the case for companies such as the video game publisher Electronic Arts Inc., which intends to update the CLV "daily for a billion customers around the world" (Driscoll, 2017). Despite that they do not require full purchase histories but only customer-level summary statistics for parameter estimation, extending such analyses to a firm's complete portfolio of customer cohorts makes the empirical application of complex models such as the Pareto/GGG an extremely computationally demanding task.

Against the background of these potential operational difficulties in real-world settings, the purpose of this research is twofold: First, this paper develops probabilistic customer behavior models capable of leveraging regularity without any noteworthy loss of sufficiency compared to their base models. In fact, the newly introduced models require exactly the same pieces of information per customer as the Pareto/GGG does, but they entail a considerably lower computational cost. Thus, the availability of such models should enable a wider range of marketing analysts and companies to benefit from improved prediction accuracy in the presence of timing regularity. Second, we evaluate the achievable theoretical improvement offered by these models relative to their baseline models and the Pareto/GGG using simulation studies and provide empirical evidence for gaining a significant increase in predictive power in a number of real-world cases. As our benchmarking results show, the performance gains are close to or on par with the Pareto/GGG, despite that the newly introduced models do not account for customer heterogeneity in timing regularity but are much easier in use.

The paper is organized as follows: In the next section, we review some recent developments in the field of probabilistic customer models after the introduction of the seminal Pareto/NBD model (Schmittlein et al., 1987), revisit the motivation to incorporate regularity and discuss relationships with the issue of clumpiness. Next, we develop a new model class for customer-base analysis, which builds on two well-established models but generalizes their transaction process component to allow for regularity within the timing patterns. The latter is achieved by replacing the Poisson assumption for interpurchase times with a more general Erlang-k distribution. We derive closed-form key expressions for fast and efficient computation. In Section 4, we report the results of two extensive simulation studies that shed light on the performance of the newly introduced model class compared with their base models and the more flexible Pareto/GGG. Furthermore, we confirm the models' capability to increase predictive performance in real-world settings using six empirical data sets from a variety of industries and provide a cost-benefit analysis vis-à-vis the more flexible Pareto/GGG. Finally, we summarize our research, derive implications and recommendations for future model users and address further research opportunities.

2. Variations of the Pareto/NBD model

Over the past decades, numerous stochastic models have been proposed to capture defection processes in noncontractual settings (Fader & Hardie, 2009). These "buy-till-you-die" models share the core assumption that an underlying, latent behavioral process triggers the current status (i.e., active vs. defected) of a customer-firm relationship. To make inferences about this status, they

[1] The Pareto/GGG reference implementation provided in the R-package BTYDplus available from https://cran.r-project.org/ on a MacBook Pro 2.2 i7 and 16GB RAM is used.

condense past transaction records into a set of observable indicator variables such as R and F (which are sufficient statistics given the assumptions of the Pareto/NBD and related models), or any other directly observable activity. All these models assume some behavioral "story" about the purchase and the defection process. In a continuous-time setting (which is the focus here), the prominent probability model developed by Schmittlein et al. (1987) proposes that an active customer buys according to a heterogeneous transaction process represented by a gamma-Poisson mixture, which results in a negative binomial distribution (NBD). Combining this with a gamma mixture of exponentials (or Pareto Type II) as a timing model for defection leads to the Pareto/NBD model.

Since then, various modifications of the Pareto/NBD have been proposed: Fader et al. (2005a) impose a different "customer death story" by restricting the occurrence of a dropout to the repurchase incidents using a (mathematically more convenient) beta-geometric (BG) dropout process. Batislam et al. (2007) and Hoppe and Wagner (2007) modified the BG/NBD (thus, MBG/NBD) by adding an additional dropout opportunity at time zero to avoid assuming zero-repeat buyers to stay alive. Jerath et al. (2011) propose and extensively discuss the effect of specifying latent attrition in discretized calendar instead of transaction time. Bemmaor and Glady (2012) generalize the Pareto by using a gamma mixture of Gompertz distributions, which allows for a non-constant hazard rate.

2.1. Prior research on incorporating timing regularity in repeat purchase models

All the above-mentioned variations of the original Pareto/NBD model are concerned with modifications of the customer dropout process, which is by definition unobservable in noncontractual settings. For the (observable) per-period purchase frequency distribution, they all share the assumption of following a Poisson process, which is equivalent to assuming independently and exponentially distributed interevent times. Because the mode of the exponential distribution is at time zero (i.e., purchases are most likely to be repeated immediately) and independence of the elapsed times imply a memoryless process, the Poisson assumption is not capable of accounting for more (or less) regular than random interpurchase timing (cf. Chatfield & Goodhardt, 1973; Schmittlein & Morrison, 1983). Alternative approaches to accommodate such duration dependence of interevent waiting times are the gamma, the generalized gamma, the Weibull and the lognormal distributions (Allenby et al., 1999; McShane et al., 2008; Winkelmann, 1995). Other attempts to allow for more flexibility and better data fit include the two-parameter COM-Poisson generalization, which can accommodate a wide range of over and under-dispersion (Mzoughia et al., 2018; Mzoughia & Limam, 2014).

In the marketing literature, the majority of contributions seeking to characterize more regular than Poisson purchases adopt the Erlang-k family of distributions. Introduced by Herniter (1971) to model interpurchase times, the Erlang-k distribution is a special case of the gamma distribution with a positive integer k as the shape parameter. Fig. 1 illustrates for a series of Erlang-k distributions the property of a non-zero mode for $k \geq 1$, with higher k values implying stronger degrees of regularity in the interevent timing patterns. The latter effect also becomes apparent from inspecting the derived random samples of corresponding timing patterns (the horizontal lines represent the time line of a customer sample, each short vertical line being a purchase). Notice that the Erlang-k collapses to the exponential distribution if k equals 1.

Chatfield and Goodhardt (1973) adopted the idea of an Erlang-2 timing process in the repeat-buying setting, which yields a count model following a "condensed" (i.e., variance less than mean) Poisson distribution. Letting the purchase rate vary across customers according to a gamma distribution, this results in a condensed NBD (or CNBD) model. Since then, the assumption of an Erlang-2 timing and its CNBD counting model counterpart have received empirical support in various packaged consumer goods categories (see, e.g., Morrison & Schmittlein, 1981, 1988; Gupta, 1988, 1991; Wu & Chen, 2000a). Furthermore, Kim and Park (1997) propose a mixture model of the exponential and the Erlang-2 distribution with heterogeneous frequency for shopping trip intervals. Fader et al. (2004) address the issue of nonstationarity in buying rates and propose a dynamic changepoint model to improve predictions for new product sales after the period of trial purchases. In the same setting, Schweidel and Fader (2009) allow customers to evolve from an initial state with exponential purchasing to a "steady state" with a more regular Erlang-2 timing process. With the exception of Wu and Chen (2000a; 2000b) none of the above-mentioned work on Erlang-k purchase timing allows for customer dropout. Furthermore, all of these studies are more concerned with improving within calibration period data fit and only a few with predictions of the aggregated future number of transactions in a customer base. The often relatively small improvements in aggregate-level model fit achieved at the expense of a substantial increase in model complexity

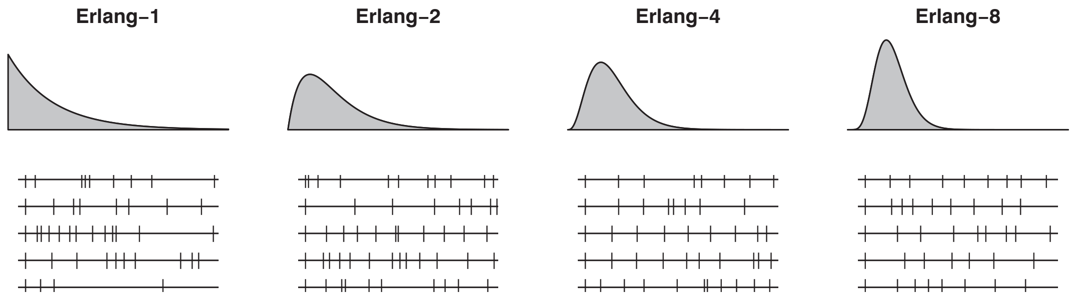

Fig. 1. Erlang-k probability density functions and sampled timing patterns.

may be one of the major reasons that many researchers in this field have concluded that moving from a Poisson to an Erlang-*k* timing model is "not worth the extra effort" (Chatfield & Goodhardt, 1973).

2.2. Combining non-Poisson purchases with latent attrition

As we will elaborate further below, it can be shown that once we move to a nonstationary setting with latent attrition (which is the typical buy-till-you-die structure), accounting for regularity in interevent timing can be leveraged to make better inference on a customer's latent defection status, which in turn translates into improved individual-level holdout predictions. Using a stochastic model with fully integrated latent attrition, Platzer and Reutterer (2016) have demonstrated that this can be achieved even for only mildly regular timing patterns, and the gain is more pronounced for stronger regularity (i.e., smaller variance in interpurchase timing). More specifically, it can be shown that the more regular historical timing patterns are, the easier it is to identify a churn event based on an overly long purchase hiatus. In other words, the benefits of accounting for regularity might be less clear if the focus is on model fit on a customer cohort level but become more evident in situations when latent attrition matters and individual-level predictions are key. Platzer and Reutterer's (2016) Pareto/GGG model generalizes the Pareto/NBD by replacing the assumption of an exponential with a flexible gamma distribution for interpurchase times. In this model, the rate parameter λ of that gamma distribution remains a measure of purchase frequency, whereas the shape parameter k captures the degree of regularity in the purchase timings. In the same way that the purchase rate λ and dropout rate μ are allowed to vary across customers, the regularity parameter k is also allowed to vary, with each of the three following independent gamma distributions (hence, the label Pareto/GGG).

Compared to prior Erlang-*k* approaches, the mixture of gammas as employed by the Pareto/GGG is extremely flexible and for values of $k < 1$ can even accommodate "less" than random interevent timings, something that Zhang et al. (2013) have termed the "clumpiness phenomenon".[2] In an extreme case, "clumpy" timing patterns exhibit a series of "clumps" or clusters with closely spaced events separated by a longer "dead-period". Such timing behavior typically emerges from switching between different activity states (e.g., a "hot" and a "cold" state with higher and lower propensities of being activated, respectively; see Schwartz et al., 2014) and frequently occurs in "binge" digital media consumption or website visitation contexts (Schweidel & Moe, 2016; Zhang et al., 2015). Park and Park (2016) demonstrate this in an online fashion retail setting and show that clumpiness seems to primarily matter on the level of customers' website visits, whereas purchases tend to follow more regular patterns. More recently, Park (2017) extends this perspective and also observes clustered visitation patterns or "shopping episodes" across multiple website visits. These findings suggest that purchases accrue within clumps of consecutive website visits followed by a longer visitation hiatus, which makes the overall interpurchase timing pattern more regular. This is in line with the applications presented by Zhang et al. (2015) and a large amount of prior work on non-Poisson interpurchase timing (see also the theoretical considerations by Kahn, 1987).

In sum, without substantial loss in data sufficiency, the Pareto/GGG combines a flexible approach to model interpurchase timing with a latent attrition process and succeeds in achieving improved individual-level predictions if there is regularity in the data. However, the flexibility and predictive accuracy delivered by the Pareto/GGG came at the cost of a significant increase in implementation complexity and in computational time and costs. This is because one can no longer derive closed-form expressions but instead has to rely on computationally intense Markov-Chain-Monte-Carlo (MCMC) simulations to make inferences based on the observed data. Efficient calculation of individual-level point estimates of future customer behavior, however, is a prerequisite to achieving scalability for firms with large customer bases that wish to integrate customer-base analysis into their ongoing operations.

3. Model development

The objective of this section is to develop a stochastic modeling approach for predicting individual-level buyer behavior, which allows marketing analysts to leverage purchase regularities "the easy way." Thus, we relax the assumption of a memoryless interpurchase timing process associated with the NBD repurchase model in the spirit of the Pareto/GGG, but aim to retain closed-form expressions for the resulting model likelihood functions that are tractable and key expressions of managerial interest that are easy to derive. In doing so, we employ the well-established BG/NBD model of Fader et al. (2005a) as a point of departure but also generalize its modified variant, the MBG/NBD (Batislam et al., 2007; Hoppe & Wagner, 2007).

In an analogous manner to the BG/NBD being able to retain predictive accuracy while reducing model complexity relative to the Pareto/NBD (see Fader et al., 2005a), our intention in employing these model assumptions is to achieve similar results with the newly developed model compared to the more complex Pareto/GGG. We expect to accomplish this with the following variations: First, just like the BG/NBD we also restrict defection events to occur only immediately after a purchase occurs. Second, rather than assuming gamma-distributed interpurchase times, we limit ourselves to the class of Erlang-*k* distributions, a subclass of the former. Third, we do not allow for heterogeneity in regularity across customers but instead assume a fixed degree of regularity for all customers within the cohort. The latter sacrifices the ability to leverage heterogeneity therein. However, for common empirical scenarios with relatively few observed purchases per customer, such a limitation should not result in significantly different estimates from those of the Pareto/GGG.

[2] Platzer and Reutterer (2016) discuss in detail the relationship between clumpiness and regularity and show that they are just opposite cases of the same underlying phenomenon, namely deviations from the Poisson assumption of exponentially distributed interarrival times.

3.1. Assumptions

The following model specification is a generalization of the BG/NBD, which relaxes the assumption of Poisson purchases by replacing the exponential with a more flexible Erlang-k distribution for interpurchase times. We denote the resulting model BG/CNBD-k, which is based on the following assumptions:

1. While active, the purchases of customers occur with Erlang-k-distributed waiting times, with rate parameter λ and integer-valued regularity parameter k.
2. Heterogeneity in λ follows a gamma distribution with shape parameter r and rate parameter α across customers.
3. Directly after each recurring purchase, there is a constant probability p that a customer becomes inactive.
4. Heterogeneity in p follows a beta distribution with parameters a and b across customers.
5. The transaction rate λ and the dropout probability p are distributed independently of one another.
6. The observation period of each individual customer begins with a purchase at time zero.

Notice that assumption 3 implies a customer dropout following a (shifted) geometric distribution, which together with assumption 4 results in the BG dropout process. By slightly modifying assumption 3 to allow for another initial dropout opportunity (with probability p) at time zero, we arrive at the MBG/CNBD-k. Below, we provide key expressions for both (M)BG/CNBD-k variants by making the same generalization of the (re)purchase process.

Assumptions 1 and 2 characterize the previously discussed CNBD (Chatfield & Goodhardt, 1973). Recall that for the special case of an estimated $k = 1$, the herein specified models are identical to their underlying base models (i.e., the BG/NBD or the MBG/NBD). However, for $k > 1$, the timing process is no longer memoryless. Compared to NBD-based models, this is also the reason why our approach requires assumption 6 to handle the first waiting time for a customer in the same way as all subsequent waiting times.[3]

It is important to keep in mind that moving away from a memoryless process not only affects the timing model but also has a number of consequences for the counting distribution of the repurchases. As already briefly mentioned in the review section, the Erlang-k timing distribution implies underdispersion in the corresponding count model (i.e., smaller variance than the mean for all values of λ), which for the special case of $k = 2$ results in what Chatfield and Goodhardt (1973) termed a condensed Poisson distribution. Furthermore, because the sum of k independently distributed exponential variables follows an Erlang-k distribution, we can interpret the purchase process as a censored Poisson process, where only every kth event is being counted (e.g., Chatfield & Goodhardt, 1973; Morrison & Schmittlein, 1988). As will be shown below, this relationship plays a key role in making the derivations of closed-form expressions feasible (see also Appendix D).

However, there is also a conceptual difference between the approach presented here and the CNBD model: The condensed Poisson is an asynchronous counting distribution, i.e., the beginning of the time interval for the counted number of repurchases is at an arbitrary point in time. In contrast, assumption 6 makes the counting process start at a known origin (i.e., when the customer was "born" at her first-ever purchase with the company). Thus, we have a synchronous counting distribution, which eases deriving the likelihood function by allowing us to treat every interpurchase event equally. Note that this implies that once we "turn off" the latent attrition process (e.g., by pushing the (M)BG parameters to the limit), for the case of $k = 2$, our approach does not collapse to the asynchronous CNBD model but to a synchronous variant thereof.[4]

3.2. Likelihood functions and parameter estimation

Given transaction history $(t_1, t_2, \ldots, t_x)$ and total observation period T, the likelihood of purchase frequency parameter λ and dropout probability p for an individual customer can be derived by multiplying the following components: (i) the likelihood of observing the waiting times $(t_1 - t_0, t_2 - t_1, \ldots, t_x - t_{x-1})$, (ii) the probability of having survived at each of the first x recurring transactions $(t_1, \ldots, t_{x-1})$, and (iii) the probability of observing no transaction within $(t_x, T]$. The hiatus at the end of the observation period can either stem from the customer defecting at the last observed transaction t_x or from encountering a waiting period longer than $T - t_x$.

For the BG/CNBD-k model variant, assumption 3 implies that there is no dropout opportunity at time $t_0 = 0$. Thus, customers with no observed recurring transactions ($x = 0$) are assumed to be alive, and the likelihood is simply the probability of observing no transactions during time T: $L(k, \lambda, p|T, x = 0) = P(X(T) = 0|k, \lambda)$.

For a customer with recurring transactions, combining the product of the three components (i) to (iii) yields

$$L(k,\lambda,p|t_1,\ldots,t_x,T,x>0) = f_\Gamma(t_1 - t_0|k,\lambda)\cdot(1-p)f_\Gamma(t_2-t_1|k,\lambda)\cdots(1-p)f_\Gamma(t_x-t_{x-1}|k,\lambda)\cdot(p+(1-p)P(X(T-t_x)=0|k,\lambda)),$$

with

$$f_\Gamma(t|k,\lambda) = \lambda^k t^{k-1} e^{-\lambda t}/(k-1)!$$

[3] In practice, compliance with this assumption is usually easy by defining time zero for each customer to be the timing of her first recorded purchase transaction.
[4] We are very grateful to the area editor of this article for highlighting this important point.

being the probability distribution function (pdf) of the Erlang-k (i.e., gamma with integer k shape parameter) distribution and

$$P(X(t)=x|k,\lambda)=\sum_{j=0}^{k-1}P_{Poisson}(X(t)=kx+j|\lambda)$$

the pdf of a synchronous counting process, which begins with an event at time zero and has Erlang-k-distributed waiting times. This last expression holds true because the Erlang-k distribution can be represented as the sum of k independent exponential variables, and thus, the corresponding counting process is equivalent to counting every kth event of a Poisson process. As already noted earlier, we need this expression to make the counting distribution synchronous. The pdf of the Poisson process is

$$P_{\text{Poisson}}(X(t)=x|\lambda)=\frac{\lambda^x t^x e^{-\lambda t}}{x!}$$

Using these results yields the following expressions for the individual likelihood function:

$$L(k,\lambda,p|t_1,\ldots,t_x,T)=\tilde{t}\cdot\left(\delta_{x>0}\,p(1-p)^{x-1}\lambda^{kx}+(1-p)^x\lambda^{kx}e^{-\lambda T}\sum_{j=0}^{k-1}\frac{\lambda^j(T-t_x)^j}{j!}\right) \tag{1}$$

with $\tilde{t}$ defined as

$$\tilde{t}:=\left(\frac{1}{(k-1)!}\right)^x\prod_{i=0}^{x-1}(t_{i+1}-t_i)^{k-1} \tag{2}$$

and $\delta_{x>0}=1$ if $x>0$, 0 otherwise.

As we already noted for the base models (BG/NBD and MBG/NBD), there is no information on the exact transaction timing ($t_1,..,t_x$) necessary for estimating the model parameters, but frequency and recency (x,t_x,T) are required as a sufficient summary of each customer's transaction history. This also applies to our presented generalized model class, which requires just one additional summary statistic to account for the timing process. To compute the log-likelihood for a given dataset, we need to maintain only the sum over the logarithmic interpurchase times as part of the logarithmic version of the above expression (2) for $\tilde{t}$:

$$\log\tilde{t}=x\cdot\log\left(\frac{1}{(k-1)!}\right)+(k-1)\sum_{i=0}^{x-1}\log(t_{i+1}-t_i)$$

Notice that this is identical to the additional data requirement imposed by the Pareto/GGG model. In expression (1), our conditioning on the timing process of the complete transaction history ($t_1,\ldots,t_x$) emphasizes that we are actually dealing with a synchronous count process, but we keep in mind that for parameter estimation the additional summary statistic on interpurchase timings is sufficient.

The derivations for MBG/CNBD-k are analogous to those above for the BG/CNBD-k, except that we have one additional dropout opportunity at time t_0, and we therefore no longer need to distinguish the cases with non-zero repurchases $x>0$ from the "zero class" of non-repeat buyers $x=0$. Thus, the individual-level likelihood function for the MBG/CNBD-k is

$$L(k,\lambda,p|t_1,\ldots,t_x,T)=\tilde{t}\cdot\left(p(1-p)^x\lambda^{kx}e^{-\lambda t_x}+(1-p)^{x+1}\lambda^{kx}e^{-\lambda T}\sum_{j=0}^{k-1}\frac{\lambda^j(T-t_x)^j}{j!}\right) \tag{3}$$

To estimate the parameters that determine the unobserved transaction rate λ and dropout probability p in the above likelihood functions, we aim to derive the equivalent expressions for a randomly chosen customer. These aggregate-level likelihood functions can be stated as closed-form expressions, which allow us to compute point estimates for the five aggregate-level model parameters (k, r, α, a, b) via maximum likelihood estimation (MLE). In practice, this is done by maximizing the logarithm of the following likelihood functions (see Appendix A for their derivations) across r, α, a, b for increasing integer-valued regularity parameter k:

$$\begin{aligned}L_{\text{BG/CNBD}-k}(k,r,\alpha,a,b|t_1,\ldots,t_x,T)=&\ \delta_{x>0}\,\tilde{t}\,\frac{\mathrm{B}(a+1,b+x-1)}{\mathrm{B}(a,b)}\frac{\alpha^r}{(\alpha+t_x)^{r+kx}}\frac{\Gamma(r+kx)}{\Gamma(r)}\\&+\tilde{t}\,\frac{\mathrm{B}(a,b+x)}{\mathrm{B}(a,b)}\left(\sum_{j=0}^{k-1}\frac{(T-t_x)^j}{j!}\frac{\alpha^r}{(\alpha+T)^{r+kx+j}}\frac{\Gamma(r+kx+j)}{\Gamma(r)}\right)\end{aligned} \tag{4}$$

$$L_{\mathrm{MBG/CNBD}-k}(k,r,\alpha,a,b|t_1,\ldots,t_x,T) = \tilde{t}\,\frac{\mathrm{B}(a+1,b+x)}{\mathrm{B}(a,b)}\frac{\alpha^r}{(\alpha+t_x)^{r+kx}}\frac{\Gamma(r+kx)}{\Gamma(r)} + \tilde{t}\,\frac{\mathrm{B}(a,b+x+1)}{\mathrm{B}(a,b)}\left(\sum_{j=0}^{k-1}\frac{(T-t_x)^j}{j!}\frac{\alpha^r}{(\alpha+T)^{r+kx+j}}\frac{\Gamma(r+kx+j)}{\Gamma(r)}\right) \tag{5}$$

The search for an optimal k can be stopped early, if consecutively higher ks do not yield higher likelihood values.

3.3. Key expressions

Once the model parameters are estimated, we can calculate various quantities of managerial interest in customer-base analysis (again, we refer to the Appendix for their derivations).

The conditional probability for a customer from the cohort to be still alive at time T is

$$P_{\mathrm{BG/CNBD}-k}(\text{alive at } T|t_1,\ldots,t_x,T,k,r,\alpha,a,b) = \left(1+\delta_{x>0}\frac{a}{b+x-1}\left(\frac{\alpha+T}{\alpha+t_x}\right)^{r+kx} \Big/ \sum_{j=0}^{k-1}\frac{(T-t_x)^j}{j!(\alpha+T)^j}\frac{\Gamma(r+kx+j)}{\Gamma(r)}\right)^{-1} \tag{6}$$

$$P_{\mathrm{MBG/CNBD}-k}(\text{alive at } T|t_1,\ldots,t_x,T,k,r,\alpha,a,b) = \left(1+\frac{a}{b+x}\left(\frac{\alpha+T}{\alpha+t_x}\right)^{r+kx} \Big/ \sum_{j=0}^{k-1}\frac{(T-t_x)^j}{j!(\alpha+T)^j}\frac{\Gamma(r+kx+j)}{\Gamma(r)}\right)^{-1} \tag{7}$$

The unconditional probability distribution of the purchase frequencies can be derived as

$$P_{\mathrm{BG/CNBD}-k}(X(t)=x|k,r,\alpha,a,b) = \frac{\mathrm{B}(a,b+x)}{\mathrm{B}(a,b)}\sum\nolimits_{j=kx}^{kx+k-1}\frac{t^j}{j!}\frac{\alpha^r}{(\alpha+t)^{r+j}}\frac{\Gamma(r+j)}{\Gamma(r)} + \delta_{x>0}\,\frac{\mathrm{B}(a+1,b+x-1)}{\mathrm{B}(a,b)}\left(1-\sum_{j=0}^{kx+k-1}\frac{t^j}{j!}\frac{\alpha^r}{(\alpha+t)^{r+j}}\frac{\Gamma(r+j)}{\Gamma(r)}\right) \tag{8}$$

$$P_{\mathrm{MBG/CNBD}-k}(X(t)=x|k,r,\alpha,a,b) = \frac{\mathrm{B}(a,b+x+1)}{\mathrm{B}(a,b)}\sum\nolimits_{j=kx}^{kx+k-1}\frac{t^j}{j!}\frac{\alpha^r}{(\alpha+t)^{r+j}}\frac{\Gamma(r+j)}{\Gamma(r)} + \frac{\mathrm{B}(a+1,b+x)}{\mathrm{B}(a,b)}\left(1-\sum_{j=0}^{kx+k-1}\frac{t^j}{j!}\frac{\alpha^r}{(\alpha+t)^{r+j}}\frac{\Gamma(r+j)}{\Gamma(r)}\right) \tag{9}$$

Deriving closed-form expressions for the *unconditional* expectations appears to be intractable. However, because the expectations are by definition the probability-weighted average of all possible outcomes, we take advantage of this relationship and compute

$$E_{\mathrm{(M)BG/CNBD}-k}(X(t),k,r,\alpha,a,b) = \sum_{x=0}^{\infty} x\cdot P_{\mathrm{(M)BG/CNBD}-k}(X(t)=x\,,k,r,\alpha,a,b). \tag{10}$$

Deriving an exact closed-form expression for the *conditional* expected number of transactions for a specific customer with purchase history (t_1, …, t_x, T) during some holdout period $(T,T+t]$, i.e., $E(X(T,T+t)|t_1,\ldots,t_x,T,k,r,\alpha,a,b)$ remains unfortunately unsolved. Schmittlein and Morrison (1983, appendix A.2) provide a solution for the conditional expectations for the asynchronous CNBD (i.e., $k=2$) counting model, but the complexity of their derivation already indicates the difficulty of any endeavor to expand this for arbitrary k with synchronous counting in combination with a latent attrition process. To keep conditional forecasts analytically tractable, we thus opt to develop an approximation of the true conditional expectation. In doing so, we can exploit the fact that counting events with Erlang-k waiting times is equivalent to counting every kth event of a Poisson process and show in Appendix D that the asynchronous counting process for Erlang-k waiting times has an expectation of $E(X(t)=j|\lambda,k)=\lambda t/k$. Given this, we propose to simply use the conditional expectation for an alive customer of the Poisson base models (see Fader et al. 2005a; Hoppe & Wagner, 2007; Wagner & Hoppe, 2008) but adjust the estimated rate parameter α by factor k. This works reasonably well but does not account for the purchase hiatus that has already occurred just prior to the holdout period and thus results in a slight underforecast bias.[5] Therefore, we refine the approximation by scaling our estimates based on the total expected transactions for the holdout period. The conditional expectation of a customer is then the probability of the customer being alive multiplied with the approximated conditional expectations of an alive customer. Recall that the motivation for leveraging regularity came from improved inferences about the latent defection status, i.e., P(alive), and for that expression we do have

[5] In a simulation study, we find this bias to be approximately 2% to 5% for common parameter settings.

closed-form solutions. As we will demonstrate in the simulation study in the next section, the approximations work sufficiently well to outperform the underlying base models in almost all of the common parameter scenarios.

Putting these pieces together, our approximate solution for the conditional expected number of transactions is as follows:

$$E_{(\mathrm{M})\mathrm{BG/CNBD}-k}(X(T,T+t)\ ,\ t_1,\ldots,t_x,T,k,r,\alpha,a,b) \approx \tilde{C}\cdot\tilde{E}_{(\mathrm{M})\mathrm{BG/CNBD}-k}(X(T,,T+t)|\text{alive at }T,k,r,\alpha,a,b)\cdot P_{(\mathrm{M})\mathrm{BG/CNBD}-k}(\text{alive at }T\mid t_1,\ldots,t_x,T,k,r,\alpha,a,b) \quad (11)$$

with $\tilde{E}(.)$ being the aforementioned approximation for the expected transactions of an alive customer during the holdout period, and $\tilde{C}$ being a bias correction factor. For the two proposed model variants, we obtain the following approximations:

$$\tilde{E}_{\mathrm{BG/CNBD}-k}(X(T,T+t)|\text{alive at }T,k,r,\alpha,a,b) := \frac{a+b+x-1}{a-1}\cdot G(r+x,k\alpha+T,a,b+x-1) \quad (12)$$

$$\tilde{E}_{\mathrm{MBG/CNBD}-k}(X(T,T+t),\text{alive at }T,k,r,\alpha,a,b) := \frac{a+b+x}{a-1}\cdot G(r+x,k\alpha+T,a,b+x) \quad (13)$$

with

$$G(r',\alpha',a',b') := 1-\left(\frac{\alpha'}{\alpha\prime+t}\right)^{r\prime}{}_2F_1\left(r',b'+1;a'+b';\frac{t}{\alpha'+t}\right) \quad (14)$$

where ${}_2F_1$ is the Gaussian hypergeometric function (see Fader et al., 2005a). The bias correction factor $\tilde{C}$ is defined such that the sum of the approximated individual-level conditional expectations across customers will match the exactly derived expected number of transactions during the holdout period at an aggregate level (see expressions (10) above), i.e.,

$$\begin{aligned}\tilde{C} := &\textstyle\sum_{i=1}^{n}(E(X(T_i+t)|k,r,\alpha,a,b)-E(X(T_i)|k,r,\alpha,a,b)) \\ &/\textstyle\sum_{i=1}^{n}\left(\tilde{E}(X(T,T+t)|\text{alive at }T,k,r,\alpha,a,b)\cdot P\left(\text{alive at }T_i\mid t_{i,1},\ldots,t_{i,x},T_i,k,r,\alpha,a,b\right)\right)\end{aligned} \quad (15)$$

4. Simulations

The just developed (M)BG/CNBD-k model family allows to account for various degrees of timing regularity in the transaction flows of a customer cohort. However, we do not yet know whether and to what extent this translates into reduced forecasting error. Thus, we need to gain a thorough understanding of the potential increase in predictive accuracy when we move from the NBD-type repurchase process to its synchronous CNBD-k variant. A natural benchmark for assessing this potential is to see how much worse the standard (M)BG/NBD model performs if there is indeed regularity in the transaction data, other things being equal. Furthermore, in order to get a sense of the performance relative to the more flexible Pareto/GGG we also benchmark the newly introduced model using data scenarios generated according to Pareto/GGG assumptions. We do this by conducting two extensive simulation studies.

4.1. Simulation design and error metrics

In our first study, the simulation setup for comparing the NBD-type to CNBD-k variants follows the full factorial designs presented by Fader et al. (2005a) and Platzer and Reutterer (2016). Specifically, we generate a large number of simulated customer cohorts by manipulating the shape and scale parameters of the mixture distributions as well as the degree of interevent timing regularity underlying the synthetic transaction histories. The parameter space is chosen such that it includes most of the commonly encountered real-world scenarios. In total, we define three levels for each of the four heterogeneity parameters ($r \in \{0.25, 0.5, 0.75\}$, $\alpha \in \{5/k, 10/k, 15/k\}$, $a \in \{0.5, 0.75, 1.0\}$, $b \in \{2.5, 5, 10\}$), combine these with four degrees of regularity ($k \in \{1, 2, 3, 4\}$), and repeat each cross combination 5 times, thus resulting in a total of $3 \times 3 \times 3 \times 3 \times 4 \times 5 = 1620$ simulated worlds. Each of these worlds consists of 4000 customers simulated over a 52-week calibration period and a 52-week holdout period.[6] These scenarios cover a wide range of settings. The average number of repeat purchases during the calibration period ranges from 0.45 to 6.22, with the share of customers with zero repeat purchases ranging from 14% to 76%. Note that we rescale α by the parameter k, to ensure that the simulated worlds result in similar purchase frequencies across the four assumed degrees of

[6] We also simulated shorter holdout periods of 4 and 16 weeks, as well as MBG/NBD vs. MBG/CNBD-k scenarios, but such modifications did not alter the key findings presented below. For simulating data and estimating the (M)BG/CNBD-k model parameters we use the model implementations provided by the R-package BTYDplus available via https://cran.r-project.org/.

regularity.

For each of these simulated worlds, we then fit both the BG/CNBD-k and the BG/NBD model to the calibration period data, calculate the conditional expected number of transactions for each customer for the holdout period ($\hat{x}_i^*$), and then assess the predictive accuracy given the "actual" number of transactions (x_i^*). This is done at both the aggregate level, by comparing the overall bias in the estimates, and the individual level, where we calculate the relative improvement in mean absolute error (MAE and MAE Lift; see also Schwartz et al., 2014):

$$\text{MAE} = \frac{1}{N}\sum_{i=0}^{N}|x_i^* - \hat{x}_i^*| \tag{16}$$

$$\text{MAE Lift} = 1 - \frac{\text{MAE}_{BG/CNBD-k}}{\text{MAE}_{BG/NBD}} \tag{17}$$

$$BIAS = \left(\sum_{i=0}^{N}\hat{x}_i^* / \sum_{i=0}^{N} x_i^*\right) - 1 \tag{18}$$

Notice that BIAS as defined in Eq. (19) is an aggregate-level measure of forecasting error and gives the percentage of over- or underforecasting. Once computed for the (here, 52-week) holdout period across the complete customer cohort, the BIAS is equivalent to the MAPE error metric (as used, e.g., by Fader et al. (2005a)), but preserving the direction of the error.

4.2. Simulation results: BG/CNBD-k versus BG/NBD

Before studying the results across all worlds, let us provide some intuition by examining in greater detail an exemplarily selected simulated cohort in Fig. 2 ($k = 2$, $r = 0.5$, $\alpha = 10/2$, $a = 0.75$, $b = 2.5$). The chart displays the actual weekly recurring purchases (solid dark line) compared to the expected aggregated numbers (dashed lines) estimated by the BG/CNBD-k and the BG/NBD models. Whereas the BG/CNBD-k nicely captures the dynamics, the BG/NBD is not able to fit the calibration period very well, which leads to overly optimistic holdout predictions. While the BG/CNBD-k comes close to identifying the true expected dropout probability $E(p) = a/(a + b) = 0.23$, the BG/NBD drastically underestimates it at only 0.15.

This observation is also reflected in Table 1 by the parameter estimates for a few selected modifications of the scenario underlying Fig. 2 (i.e., $k \in \{2,3\}$, $r = 0.5$, $\alpha/k \in \{5/k, 10/k\}$, $a = 0.75$, $b \in \{2.5,5\}$). Overall, we find that the BG/CNBD-k estimates recover the underlying data-generating parameters very well.

However, the BG/NBD estimates are negatively affected by regularity and the deviations from the true values are stronger for the dropout rate (a and b) than for the purchase rate parameters (r and α). Take for instance the scenario $k = 2$, $r = 0.5$, α/k=5,

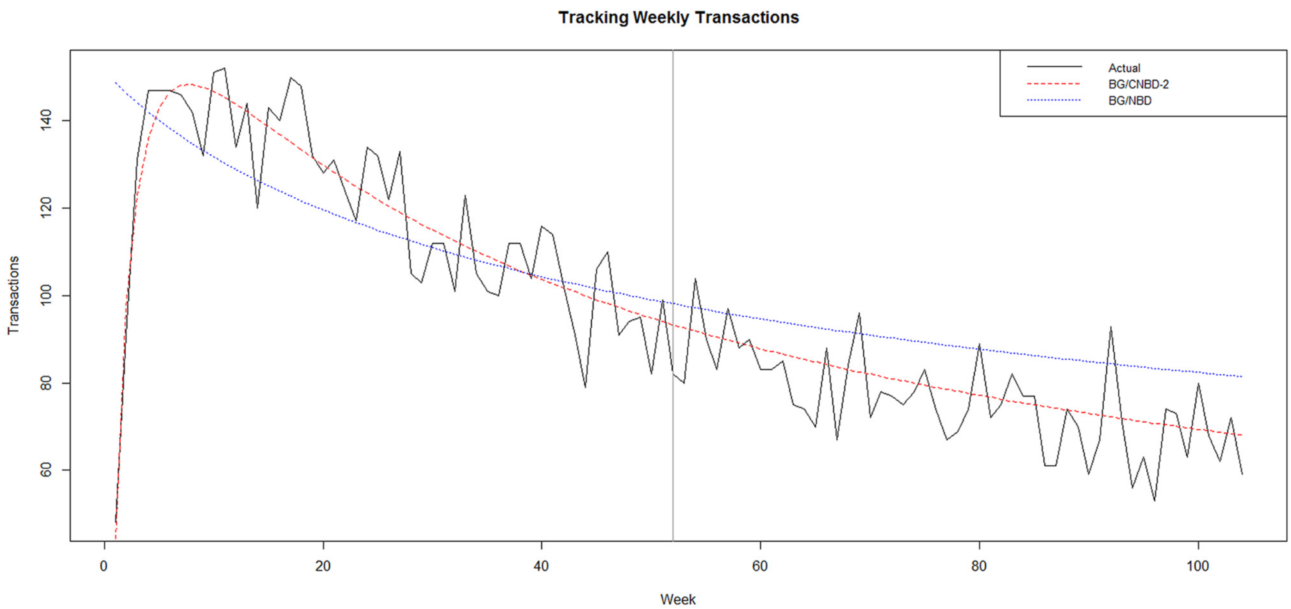

Fig. 2. Comparison of estimated versus actually recurring weekly transactions.

Table 1
Recovery of true parameter values for eight selected scenarios.

k			r			α/k			a			b		
True	NBD	CNBD-k	True	NBD	CNBD-k	True	NBD	CNBD-k	True	NBD	CNBD-k	True	NBD	CNBD-k
2	1	2	0.5	0.52	0.52	2.5	3.24	2.64	0.75	0.61	0.71	2.5	2.54	2.32
2	1	2	0.5	0.52	0.52	2.5	3.03	2.58	0.75	0.73	0.73	5	6.58	4.61
2	1	2	0.5	0.56	0.52	5	7.72	5.34	0.75	0.41	0.68	2.5	2.38	2.34
2	1	2	0.5	0.55	0.53	5	7.1	5.25	0.75	1.62	1.1	5	25.23	7.23
3	1	3	0.5	0.51	0.5	1.67	2.2	1.68	0.75	0.65	0.76	2.5	2.99	2.51
3	1	3	0.5	0.5	0.51	1.67	2.1	1.69	0.75	0.78	0.79	5	8.46	5.1
3	1	3	0.5	0.57	0.51	3.33	5.47	3.41	0.75	0.31	0.63	2.5	2	1.96
3	1	3	0.5	0.54	0.51	3.33	4.96	3.35	0.75	2.04	0.92	5	49.45	5.75

$a = 0.75$, and $b = 5$: Here, the BG/NBD estimates an a (b) which is two (five) times higher than the true values leading to a dramatic underestimation of the dropout probability. This observation indicates a serious deficiency of the NBD repurchase model in the presence of purchase timing regularity. Apparently, the NBD model falls short in leveraging the observation of an "overdue" transaction to distinguish active from inactive customers and instead interprets this as an exceptionally longer-than-average repurchase hiatus.

This finding from a few simulated scenarios with regular purchase timing patterns generalizes to other parameter settings as well. The key results are displayed in Figs. 3a-c, which show the distributions of our error measures across all simulated worlds, divided into four groups by the degree of underlying regularity, as boxplots.[7] As expected, in the case of $k = 1$, we obtain identical results for BG/NBD and BG/CNBD-k, indicating that the latter is indeed a generalization of its underlying base model. However, in the presence of even mildly regular purchase patterns ($k = 2$), we can see that applying the BG/CNBD-k model results in significantly lower aggregate and individual-level errors relative to the base model. As illustrated in the above example (see Fig. 2), the BG/NBD base model exhibits a significant overestimation bias for regular patterns. As shown in Fig. 3a for the aggregate level, the extent of this over-forecasting is, on average, approximately 10% across the simulated worlds for $k > 1$, with the median bias being 9.4% for $k = 2$ and 10.6% for $k = 4$. Further, the individual-level forecast errors depicted in Figs. 3b-3c clearly show that the higher the regularity in the data, the greater the lift in predictive accuracy when a model accounts for this regularity compared to the NBD-type baseline model. This finding makes intuitive sense, as stronger regularity helps us to more quickly identify a churned customer.

To further examine the driving forces behind these simulation outcomes, we fitted a Classification And Regression Tree (CART) model using a recursive partitioning procedure introduced by Breiman et al. (1984). More specifically, we use an implementation of a regression tree in the R package rpart (Therneau et al., 2019) which aims at maximizing the between-group variances at each split. We regressed the obtained MAE lift values on the data set characteristics implied by the simulation design parameters, i.e., regularity (k), expected purchase rate ($E(\lambda) = r/\alpha$) and expected dropout probability ($E(p) = a/(a + b)$). Fig. 4 displays the resulting tree structure which confirms the impression we already gained from inspecting the descriptives in Figs. 3b-3c. Note that the tree represents a series of splits of the simulated worlds into subgroups and each node displays the mean MAE lift together with the number of worlds represented by each branch. While the complete tree accounts for an R-squared of 0.87 the incremental improvement in explained variance is clearly decreasing from top to bottom. Unsurprisingly, the first split separates datasets with a regularity parameter of $k = 1$ (for which the predictive individual-level performance of the BG/NBD and the BG/CNBD-k is identical) from others. According to the terminal nodes at the far right hand side branch of the tree the highest MAE lift can be expected for datasets with medium to higher purchase rates ($E(\lambda) \geq 0.0337$) and higher degrees of regularity $k \geq 3$. This observation makes intuitive sense, because the model's capability to leverage regularities in inter-event timings for correctly spotting churn requires a sufficient number of observed past purchase events. For about one third of the simulated worlds we observed an MAE lift of more than 10%. But also for less regular datasets the MAE lift can be substantial when moving from the BG/NBD base model to the appropriate BG/CNBD-k model.

We also find a small share of scenarios for which the approximation of the conditional expectation comes at such a cost that it is not able to outperform the BG/NBD model.[8] These boundary cases are characterized by scenarios with a very low purchase frequency coupled with a very low dropout probability at each of these rare purchase events. In these scenarios, nearly all of the customers remain alive (>93%) at the end of the observation period, and the theoretical gain from exploiting regularity for inferring attrition becomes smaller than the error introduced by the approximation in our expression for the conditional expectations. Conversely, in a setting with even mildly regular patterns and at least 7% of customers expected to have churned the new model classes provide superior predictive performance at the individual level.

Overall, we observe larger increases in predictive accuracy at the customer level for stronger degrees of regularity. These results, as well as the magnitude of the improvement, are also closely in line with the findings from the simulation study on the Pareto/GGG reported by Platzer and Reutterer (2016).

7 Note that in all 1620 simulations, the regularity parameter k is correctly recovered from the generated data.

8 Notice that the respective terminal node displayed in Fig. 4 represents 4.4% of the scenarios. Still, half of the scenarios in this node also contains cases with slightly positive MAE lift values.

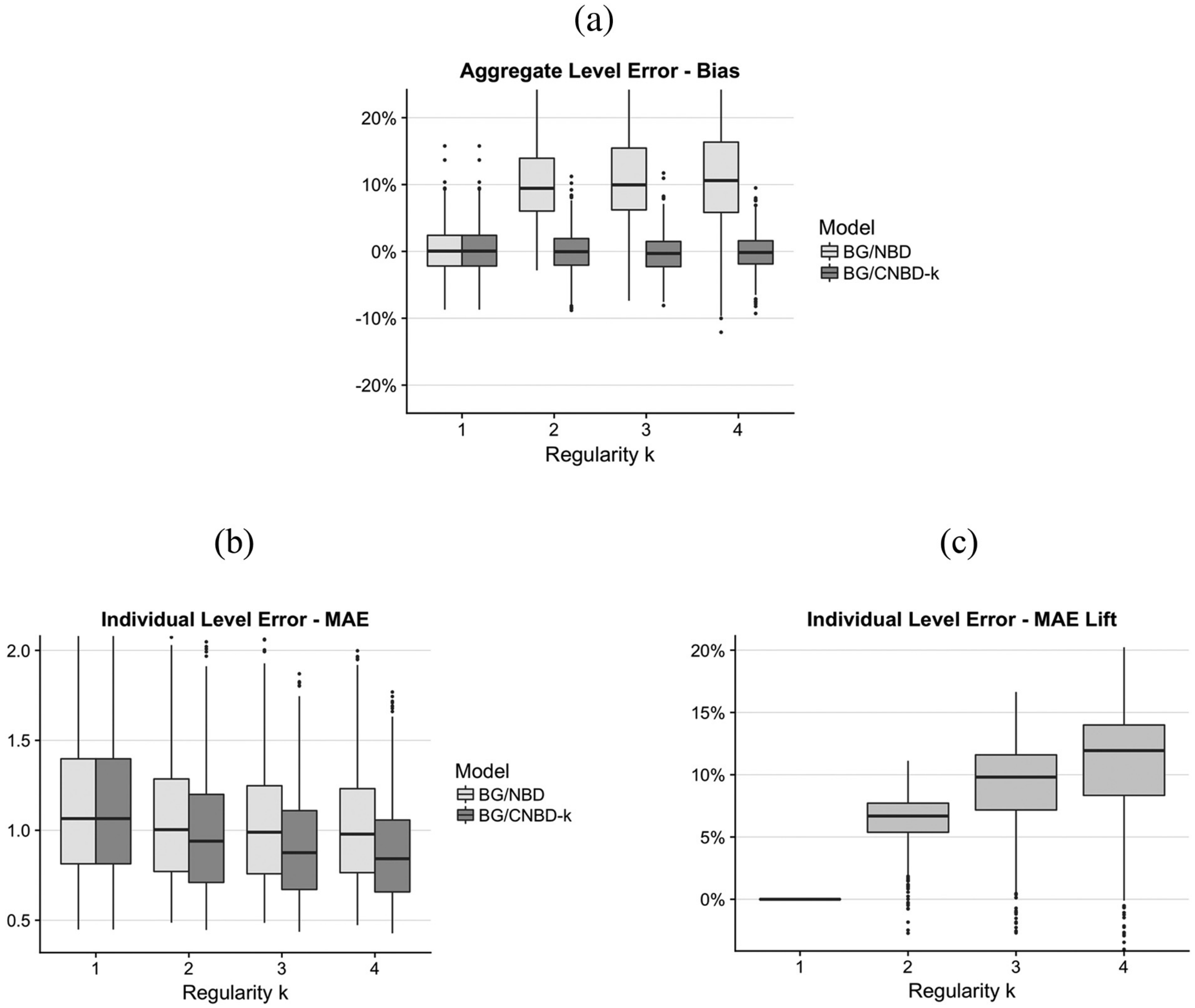

Fig. 3. Predictive errors for various degrees of regularity.

4.3. Benchmarking against the Pareto/GGG

To investigate the degree of performance loss when moving from the flexible, but computationally demanding Pareto/GGG to the BG/CNBD-*k* and to examine the conditions under which this loss is small, we perform a second simulation study. In doing so, we use exactly the same simulation setup as in the benchmark study conducted by Platzer and Reutterer (2016).[9] Similar to the first simulation, we again fit a CART model to regress the observed MAE lift values on the six simulation design variables which define the respective shape and rate parameters of the gamma distributions for the regularity parameter $k \sim Gamma(t,\gamma)$, the purchase rate $\lambda \sim Gamma(r,\alpha)$, and the dropout rate $\mu \sim Gamma(s,\beta)$.

Overall, we observe an average decrease in performance by 9% for the BG/CNBD-*k*. The resulting tree structure depicted in Fig. 5 accounts for an R-squared of 0.73 and clearly reveals that most of the individual-level performance differences between the two models is driven by the two shape parameters *s* and *t* of the dropout rate and purchase regularity distribution, respectively. As represented by the terminal node on the right hand side of Fig. 5 for 20% of all simulated worlds, scenarios characterized by combinations of lower *s* ($s < 0.5$) and higher *t* ($t \geq 7$) the BG/CNBD-*k* model is almost on par with the Pareto/GGG. Keeping in mind that the coefficient of variation of a gamma distribution is $1/\sqrt{shape}$, higher shape parameter values indicate more homogeneity in the resulting distributions. Thus, using the BG/CNBD-*k* can be expected to result in decent predictive performance for data sets with elevated heterogeneity in the dropout process and more homogeneous purchase regularity patterns. The opposite is true for scenarios with more heterogeneous purchase regular-

[9] For deriving the Pareto/GGG parameters, we made 6000 MCMC draws in 4 chains in parallel using the implementation in the R-package BTYDplus on a laptop with a 2.5GHz quad-core Intel Core i7. Using this configuration, parameter estimation requires 72 min for a simulated world with $N = 1000$ and 337 (!) minutes for $N = 4000$ customers, on average. We note that despite the parameter draws could further be parallelized, the Pareto/GGG computation times seem to grow stronger than linear with the number of customers.

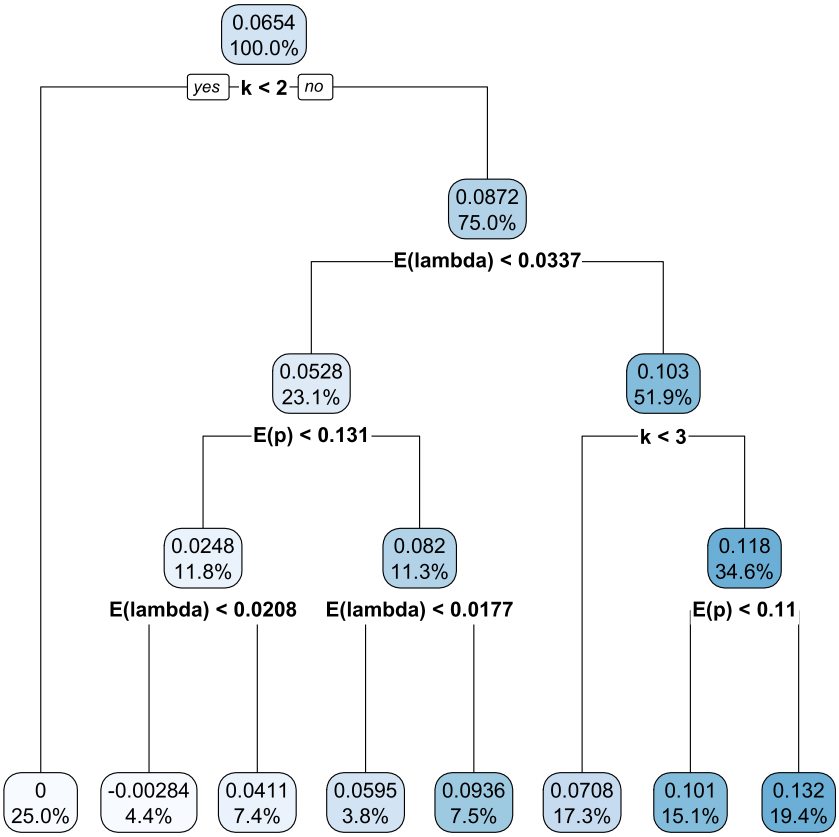

Fig. 4. Regression tree for predicting MAE lift against BG/NBD.

ity but lower heterogeneity in dropout rates as characterized by the terminal nodes on the left hand side of the tree structure. Here, the model flexibility of the Pareto/GGG allows to better leverage overly long waiting times at the end of the calibration period for predicting churn than the BG/CNBD-k does. The fact that higher heterogeneity in the regularity parameter k favors the Pareto/GGG is not surprising and in line with findings from Platzer and Reutterer (2016). However, the computation time required for estimating the BG/CNBD-k model and making customer-level predictions is over one hundred times faster than the Pareto/GGG, based on the authors' reference implementations.

5. Empirical applications

Next, we benchmark the newly introduced model classes against their base models, as well as against the Pareto/GGG, for six distinct real-world data sets. The datasets are the same as those used in Platzer and Reutterer (2016):

- CDNOW: 2357 customers of an online CD store acquired in the first quarter of 1997 and observed over 1.5 years. This is likely the most widely used benchmark dataset in testing models for customer-base analysis (see, e.g., Fader & Hardie, 2001; Fader et al., 2005a, 2005b; Batislam et al., 2007; Bemmaor & Glady, 2012; Zhang et al., 2015).
- Apparel and accessories: 831 customers of an online apparel and accessories retailer, acquired in April 2009 and observed over the course of one year (Zhang et al., 2015).
- Donations: 21166 donors to a nonprofit organization, acquired in the first half of 2002 and observed over 4.5 years (Bemmaor & Glady, 2012; Malthouse, 2009; Schweidel & Knox, 2013).
- Groceries: 1525 customers of an online grocery retailer with first recorded purchase in the first quarter of 2006, observed over two years (Platzer & Reutterer, 2016).
- Dietary supplements and office supplies: Two additional data sets, each consisting of 1000 customers with an observation period of 35 weeks (Platzer & Reutterer, 2016).

Table 2 provides an overview of the six data sets together with some key descriptive statistics. As we can see from these statistics, the datasets represent rather diverse empirical settings covering a wide range of repurchase frequencies and dropout rates. Based on the Wheat and Morrison (1990) estimator for regularity, of these six, only the CDNOW dataset follows the Poisson assumption for the purchase process decently (i.e., $k \sim 1$), four of them exhibit regular ($k > 1$) and one dataset irregular timing

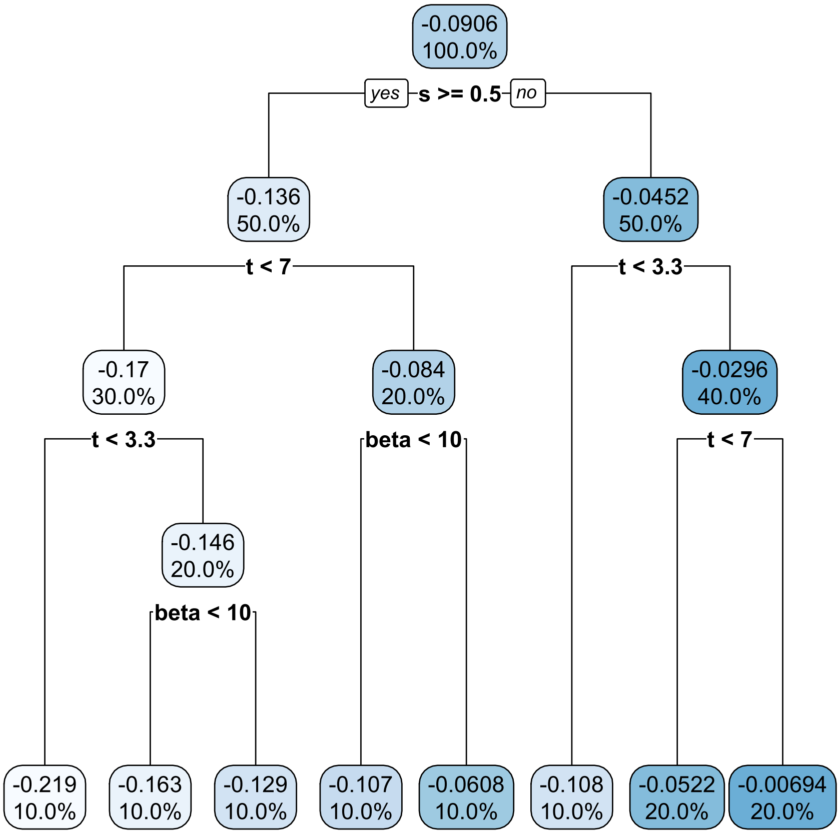

Fig. 5. Regression tree for predicting MAE lift against Pareto/GGG.

Table 2
Descriptive statistics for the six empirical data sets.

Data set	Cohort size	Period length in weeks (calibration/holdout)	Share of inactive customers (calibration/holdout)	Share of customers with 4 or more transactions (calibration/holdout)	Mean no. of purchases (calibration/holdout)	Mean no. of purchases among actives (calibration/holdout)	Wheat and Morrison (1990) estimator for regularity	Share of "clumpy" customers (Zhang et al., 2015)
CDNOW	2357	39/39	60%/71%	6%/4%	1.04/0.80	2.60/2.75	1.0	9.04%
Apparel & accessories	831	39/13	46%/81%	12%/1%	1.90/0.38	3.53/2.04	0.6	15.41%
Donations	21166	182/52	52%/81%	5%/1%	1.24/0.28	2.58/1.53	2.2	15.85%
Groceries	1525	52/52	39%/66%	27%/16%	3.65/2.22	5.96/6.56	2.5	24.62%
Dietary supplements	1000	26/9	76%/92%	1%/0%	0.40/0.09	1.70/1.10	2.0	1.72%
Office supply	1000	26/9	71%/83%	6%/1%	0.90/0.32	3.11/1.89	2.1	3.42%

pattern.[10] In the last column of Table 2 we also report the share of customers classified as "clumpy" at the 5% confidence level according to the individual-level entropy measure proposed by Zhang et al. (2015).[11] Based on this measure, almost 25% of the customers in the grocery data are characterized by clumpy repurchase patterns, whereas this share is less than 2% for customers in the dietary supplements dataset.

[10] Wheat and Morrison (1990) provide a data-set-level summary statistic, which can serve as a diagnostic check for regularity. Their regularity estimator is defined by the simple formula $\hat{k} = (1-4Var(M))/8Var(M)$, with $M \coloneqq itt_1/(itt_1 + itt_2)$ and two intertransaction times (itt_1 and itt_2) drawn from an individual's purchase history (see Wheat & Morrison, 1990, pp.88). This definition implies that we are typically excluding a large proportion of the customer base in any dataset. More precisely, we exclude those customers who made less than two repeat purchases because we do not observe the required two interpurchase times for them.

[11] We used the normalized clumpiness measure $H_p : 1 + \frac{\sum_{i=1}^{t_x+1} \log(iit_i) \cdot iit_i}{\log(t_x+1)}$ as suggested by Zhang et al. (2015).

Table 3 displays the estimated parameters for the best fitting (in terms of model likelihoods) (M)BG/CNBD-k models and compares them to their NBD-type base models. We also report the respective weekly mean interpurchase times ($\overline{itt}$) implied by the estimated parameters (i.e., $\overline{itt} = k{\cdot}\alpha/r$) rounded to the next integer value. Additionally, Table 3 includes the expected dropout probability ($E(p) = a/(a + b)$) and the maximized log-likelihood value (LL). A number of observations are noteworthy: First, the BG/NBD and the MBG/NBD already result in substantially different parameter estimates. For all data sets, the MBG/NBD estimates shorter interpurchase times and higher dropout probabilities. This is the expected effect for data sets with a significant share of non-recurring customers, as the BG/NBD assumes all inactive customers to still be alive, thus inflating the interpurchase time estimates. Furthermore, higher shares of active customers showing more transactions imply more dropout opportunities, and the estimated dropout probability at each of these opportunities is estimated to be lower.

For the two data sets with no regular timing patterns, the estimates for BG/CNBD-k and MBG/CNBD-k simply revert to those of their base models. For the other four data sets, the data suggest that Erlang-2 offers the best fit among the class of Erlang-k distributions for the observed interpurchase times. In these cases, the deviation from the Poisson assumption also significantly impacts all other parameters. Because the Erlang-2 distribution does not allow for as much variation in interpurchase times as the exponential distribution, an overly long purchase hiatus at the end of the calibration period is thus less likely to be interpreted as a regular interpurchase time and more likely to be a signal for dropout. In the case of the donations data set, the estimates

Table 3
Parameter estimates and model comparison for the empirical data sets.

Data set	BG/NBD	MBG/NBD	BG/CNBD-k	MBG/CNBD-k
CDNOW	$r = 0.24$ $\alpha = 4.41$ $a = 0.79$ $b = 2.43$ $\overline{itt} = 18w$ $E(p) = 0.25$ $LL = -9582$	$r = 0.52$ $\alpha = 6.18$ $a = 0.89$ $b = 1.61$ $\overline{itt} = 12w$ $E(p) = 0.36$ $LL = -9582$	$k = 1$, see BG/NBD	$k = 1$, see MBG/NBD
Apparel & accessories	$r = 0.32$ $\alpha = 3.7$ $a = 0.73$ $b = 3.05$ $\overline{itt} = 11w$ $E(p) = 0.19$ $LL = -5285$	$r = 0.56$ $\alpha = 4.88$ $a = 0.85$ $b = 2.64$ $\overline{itt} = 9w$ $E(p) = 0.24$ $LL = -5286$	$k = 1$, see BG/NBD	$k = 1$, see MBG/NBD
Donations	$r = 0.51$ $\alpha = 64.4$ $a = 0.67$ $b = 12.47$ $\overline{itt} = 127w$ $E(p) = 0.05$ $LL = -146901$	$r = 0.57$ $\alpha = 67.51$ $a = 0.55$ $b = 8.19$ $\overline{itt} = 118w$ $E(p) = 0.06$ $LL = -146899$	$k = 2$ $r = 0.46$ $\alpha = 21.42$ $a = 0.82$ $b = 3.85$ $\overline{itt} = 93w$ $E(p) = 0.18$ $LL = -144713$	$k = 2$ $r = 1.37$ $\alpha = 38.84$ $a = 0.53$ $b = 0.83$ $\overline{itt} = 57w$ $E(p) = 0.39$ $LL = -144541$
Groceries	$r = 0.35$ $\alpha = 2.58$ $a = 0.5$ $b = 3.64$ $\overline{itt} = 7w$ $E(p) = 0.12$ $LL = -15837$	$r = 1.26$ $\alpha = 6.04$ $a = 0.37$ $b = 0.72$ $\overline{itt} = 5w$ $E(p) = 0.34$ $LL = -15782$	$k = 2$ $r = 0.36$ $\alpha = 1.13$ $a = 0.52$ $b = 2.79$ $\overline{itt} = 6w$ $E(p) = 0.16$ $LL = -15054$	$k = 2$ $r = 1.33$ $\alpha = 2.81$ $a = 0.42$ $b = 0.79$ $\overline{itt} = 4w$ $E(p) = 0.35$ $LL = -14978$
Dietary supplements	$r = 0.3$ $\alpha = 14.27$ $a = 1.68$ $b = 11.57$ $\overline{itt} = 48w$ $E(p) = 0.13$ $LL = -1912$	$r = 0.4$ $\alpha = 15.41$ $a = 0.72$ $b = 3.1$ $\overline{itt} = 39w$ $E(p) = 0.19$ $LL = -1912$	$k = 2$ $r = 0.2$ $\alpha = 2.83$ $a = 1.9$ $b = 3.48$ $\overline{itt} = 28w$ $E(p) = 0.35$ $LL = -1911$	$k = 2$ $r = 0.72$ $\alpha = 4.62$ $a = 1.22$ $b = 0.96$ $\overline{itt} = 13w$ $E(p) = 0.56$ $LL = -1910$
Office supply	$r = 0.19$ $\alpha = 4.08$ $a = 287$ $b = 10000$ $\overline{itt} = 21w$ $E(p) = 0.03$ $LL = -2954$	$r = 0.2$ $\alpha = 4.12$ $a = 282$ $b = 10000$ $\overline{itt} = 21w$ $E(p) = 0.03$ $LL = -2954$	$k = 2$ $r = 0.18$ $\alpha = 1.34$ $a = 0.28$ $b = 1.14$ $\overline{itt} = 15w$ $E(p) = 0.20$ $LL = -2898$	$k = 2$ $r = 0.33$ $\alpha = 1.61$ $a = 0.37$ $b = 0.69$ $\overline{itt} = 10w$ $E(p) = 0.35$ $LL = -2899$

Table 4
Predictive performance comparison across models and data sets.

(a) MAE

Model:	$k = 1$		$k = 2$			
	CDNOW	Apparel	Donations	Groceries	Dietary	Office
BG/NBD	0.7855	0.4403	0.3598	1.5789	0.1676	0.3083
MBG/NBD	0.7648	0.4337	0.3585	1.5148	0.1665	0.3084
BG/CNBD-*k*	0.7855	0.4403	0.3337	1.4748	0.1503	0.2939
MBG/CNBD-*k*	0.7648	0.4337	0.3137	1.4329	0.1438	0.2882
Pareto/GGG	0.7551	0.4449	0.2937	1.3887	0.1516	0.2823

(b) MAE Lift (%)

Model	$k = 1$		$k = 2$			
	CDNOW	Apparel	Donations	Groceries	Dietary	Office
BG/NBD						
MBG/NBD	+2.6%	+1.5%	+0.4%	+4.1%	+0.7%	−0.0%
BG/CNBD-*k*	+0.0%	+0.0%	+7.3%	+6.6%	+10.4%	+4.7%
MBG/CNBD-*k*	+2.6%	+1.5%	+12.8%	+9.2%	+14.2%	+6.5%
Pareto/GGG	+3.9%	−1.0%	+18.4%	+12.0%	+9.5%	+8.4%

(c) BIAS

Model	$k = 1$		$k = 2$			
	CDNOW	Apparel	Donations	Groceries	Dietary	Office
BG/NBD	−12.1%	8.4%	19.6%	18.2%	58.2%	−0.8%
MBG/NBD	−16.2%	5.7%	19.0%	19.1%	57.5%	−0.6%
BG/CNBD-*k*	−12.1%	8.4%	6.7%	13.3%	31.0%	0.8%
MBG/CNBD-*k*	−16.2%	5.7%	3.0%	17.4%	28.8%	−3.4%
Pareto/GGG	−12.6%	13.1%	−10.9%	12.7%	47.6%	−7.7%

for the expected dropout probability increase drastically from 0.06 for the MBG/NBD to 0.39 for the MBG/CNBD-*k*, and the mean interpurchase time drops from 118 weeks to 57 weeks. In the case of the office supply data set, the base models have problems even converging, and the likelihood optimization[12] is aborted once parameter b exceeds 10,000. Note that this "blowing up" of the beta distribution shape parameter(s) implies homogeneity in p, taking on a value of $E(p) = a/(a + b)$. Once we allow for regularity, this case also yields parameter estimates that seem to be valid.

Aside from seemingly more realistic parameter estimates and better data fit, the other key evaluation criterion we are interested in is whether the "regularity-sensitive" model class also results in improvements in terms of predicting future customer behavior. Moreover, we also wish to examine how the comparatively parsimonious and simple MBG/CNBD-*k* compares to the more flexible, yet much more complex and computationally intense, Pareto/GGG model.

Table 4 presents the error metrics we have already introduced in the simulation study for all combinations of the employed models and empirical data sets. The percentage improvements in MAE are computed relative to the respective BG/NBD baseline. A couple of important observations and insights can be derived from inspecting Table 4. Based on the individual-level error metric (MAE), the MBG/NBD performs slightly better than or at least on par with the BG/NBD model in all six data cases.

Note that for the two cases that retain the Poisson assumption ($k = 1$), the predictions of the more flexible model classes are identical to those of their base models. In the presence of regularity, however, the BG/CNBD-*k* and MBG/CNBD-*k* perform significantly better than their respective base models, with the relative improvement in MAE ranging from approximately 5% to 14%. It also becomes obvious from Table 4b that the MBG/CNBD-*k* consistently performs best of the four maximum likelihood estimated models. When comparing the performance of the MBG/CNBD-*k* relative to the Pareto/GGG one interesting observation emerges from looking at the interplay between regularity and the share of clumpy customers as reported in Table 2: If regularity is present in the database, higher shares of clumpy customers imply more customer-level heterogeneity in purchase regularity. In such situations (i.e., for the donations and the groceries data sets), the Pareto/GGG can leverage this heterogeneity and outperforms the MBG/CNBD-*k* approach (which does not account for individual-level heterogeneity). The opposite is true for the datasets with lower shares of clumpy customers (i.e., dietary supplement and office supply) where the MBG/CNBD-*k* performs even better or at least almost as well as the Pareto/GGG. This is remarkable because the Pareto/GGG offers significantly greater flexibility by allowing for customer-level heterogeneity in the degree of regularity. However, as our results show, the latter can only be

[12] Using the limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm (Byrd et al., 1995).

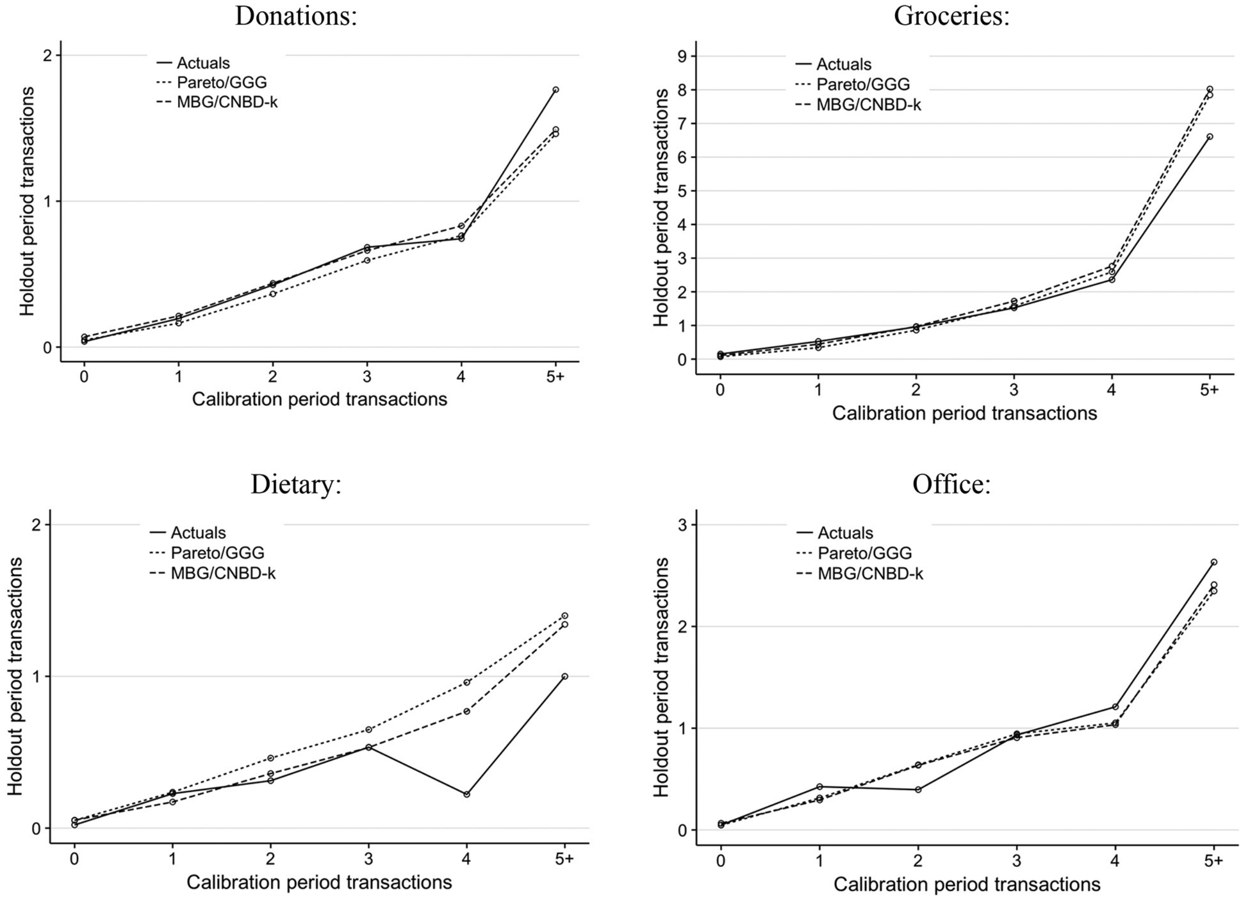

Fig. 6. Conditional expectations across data sets.

taken advantage of if (i) enough heterogeneity is present within the customer cohort, and (ii) the share of very active customers (with 5 or more transactions during the calibration period) is large enough, such that the model can also update its regularity estimate at an individual level.

Overall, the results for the aggregate error metric (see Table 4c) confirm our assessment of individual-level performance. Again, for three out of four datasets with non-Poisson purchases (here, Erlang-2), the CNBD-*k* models considerably reduce the overall forecasting biases. An exception is the office supplies data set; however, we have here very low holdout forecast errors across all models. Interestingly, except for the grocery data set, the MBG/CNBD-*k* even outperforms the costly Pareto/GGG on this dimension.

Table 5
Computational cost for parameter estimation and conditional expectations.

Data set	Cohort Size	BG/NBD	MBG/NBD	BG/CNBD-*k*	MBG/CNBD-*k*	Pareto/NBD (HB)	Pareto/GGG	Speedup (MBG/CNBD-*k* vs. Pareto/GGG)
(a) Parameter estimation (computation times in seconds)								
CDNOW	2357	0.20	0.19	0.75	0.77	18.86	565.68	733.7
Apparel	831	0.08	0.10	0.31	0.30	6.65	199.44	660.4
Donations	21166	2.72	3.01	9.04	12.26	677.31	20319.36	1657.6
Groceries	1525	0.12	0.16	0.67	0.65	12.20	366.00	562.2
Dietary	1000	0.18	0.09	0.65	0.66	8.00	240.00	363.1
Office	1000	0.36	0.39	0.92	0.84	8.00	240.00	285.4
(b) Conditional expectations (computation times in seconds)								
CDNOW	2357	0.00	0.00	0.00	0.00	75.42	113.14	56568.0
Apparel	831	0.00	0.00	0.00	0.00	26.59	39.89	39888.0
Donations	21166	0.01	0.01	0.73	0.72	677.31	1015.97	1418.9
Groceries	1525	0.00	0.00	5.23	1.41	48.80	73.20	52.0
Dietary	1000	0.00	0.00	0.11	0.12	32.00	48.00	396.7
Office	1000	0.00	0.00	0.22	0.16	32.00	48.00	309.7

The predictions for the number of transactions in the holdout periods reported in Table 4 are conditional on the number of observed transactions in the calibration period. For the CNBD-k models, these are derived using the approximation given in Eq. (11), and for the Pareto/GGG, they are based on the simulated posterior expectations for the MCMC parameter draws. In Fig. 6, we show these conditional expectations for the MBG/CNBD-k and the Pareto/GGG along with the average of the actual number of holdout-period transactions, broken down by the calibration-period number of repeat transactions.[13]

Both models provide excellent predictions for the holdout period, except for customer groups with a very large number of transactions in the calibration period. However, it is important to note that those groups are typically very small. As an extreme case, in the dietary supplements dataset, there are just 9 customers with 4 and 6 persons with 5 or more purchases in the calibration period. Here, the MBG/CNBD-k offers better predictions than the Pareto/GGG, which is also the case (but less pronounced) for most of the customer groups in the donation dataset. For the grocery and the office supplies datasets, the differences between the two models are very small. Something that is hard to spot from Fig. 6 is the relative performance for the so-called "zero class" (the typically large fraction of inactive customers in the calibration period; see Table 2 for the exact numbers). For these groups, the more flexible Pareto/GGG comes slightly closer to the actual holdout-period transaction numbers for the donations, the dietary supplement, and office supply dataset, whereas the MBG/CNBD-k is closer in the grocery dataset.

We have repeatedly referred to the increased model complexity and accordingly higher computational cost required for estimating the Pareto/GGG, but we have been relatively silent thus far on the magnitude of this extent. To gain a better understanding of the computational cost-benefit tradeoff of moving from maximum likelihood estimated models to the Pareto/GGG, Table 5 reports the computation times required to estimate model parameters (panel a) and deriving conditional expectations (panel b) for the models and data sets we used in our empirical performance comparison.[14] In addition, we also compare the computation times for a hierarchical Bayesian (HB) implementation of the Pareto/NBD (see Ma & Liu, 2007) with the Pareto/GGG. This allows us to obtain a sense of the additional computational effort required to estimate flexible, individual-level regularity parameters. It turns out that this is very expensive: using an otherwise identical MCMC sampling scheme, the additional computational effort amounts to a factor of 30.

The last column in Table 5 a/b reports the speedup in computation time (in terms of multiples) that we can gain when moving from the Pareto/GGG to the more parsimonious MBG/CNBD-k model. Regarding parameter estimation, the (M)BG/CNBD-k is approximately 300 to 1500 times (!) faster than the Pareto/GGG. For deriving the conditional expectations, this speedup is approximately 50 to 400 times for CNBD-k models with $k = 2$. In contrast, the additional cost for the CNBD-k versions compared to their NBD-counterparts is relatively modest. Because the models need to be evaluated for various values of k, parameter estimation for the former is just about 4 times slower than the latter.[15] This appears to be a "fair price" for the achievable lift in predictive holdout accuracy reported in Table 4, whereas the Pareto/GGG will reach its limits when applied to routine real-world applications in data rich environments.

6. Discussion

We have developed a new class of models that comply with a recently revived discussion on the benefits of extending the RF-framework of stochastic models for customer-base analysis by adding another individual-level metric summarizing (ir)regularities in intertransaction timing. The model variants arising from this model class are direct extensions of the well-established BG/NBD model (Fader et al., 2005a) and account for varying degrees of regularity in purchase behavior. We achieve this by replacing the memoryless exponential interarrival process of the NBD repeat-buying component with a duration-dependent Erlang-k process. The resulting model families are denoted BG/CNBD-k and MBG/CNBD-k, where k indicates the degree of regularity assumed in the timing patterns. The attrition process of the latter model class (i.e., MBG) differs from the former by allowing for another dropout option after the initial purchase (see Batislam et al., 2007; Hoppe & Wagner, 2007).

The key contributions of this research are as follows: The newly developed models offer marketing analysts the option to account in their predictions for a wide range of purchase timing regularity. If the interevent timing process follows more regular than random patterns, retaining the NBD assumption leads to an overestimation bias in the predictions. According to our simulation study, this bias, on average, is just over 10% of incorrectly predicted future transactions even for only mildly regular timing patterns. Our extensive simulation and empirical application studies also clearly illustrate that accounting adequately for regularity can yield significantly improved prediction accuracy. In particular, we find that the MAE lift can increase by 13% for higher degrees of regularity and medium to higher purchase rates; in the empirical settings we analyzed, the mean predictive errors can be reduced by up to 14% relative to the respective base model. Furthermore, the model generalizations presented in this paper nicely nest their underlying base models as special cases (i.e., for $k = 1$). Thus, we add a more general and flexible alternative to the original (M)BG/NBD models to the customer-base analyst's toolbox by extending NBD towards a CNBD-k timing model. This gain in flexibility comes at an only moderate additional cost in terms of data requirements and computation time. The data structure required to derive the necessary summary statistics (i.e., recency, frequency, and the sum of logarithmized interpurchase times) is simple and requires the same event logs as every other RF-based stochastic customer behavior model. Thus, we expect

[13] Because of space restrictions and their overall superior performance compared to the other model variants, we only exhibit the plots for these two models and for datasets with $k = 2$.

[14] For deriving the Pareto/GGG parameters, we made 8000 MCMC draws in 4 chains in parallel using the implementation in the R-package BTYDplus on a laptop with a 2.5GHz quad-core Intel Core i7 (to achieve convergence for the donations dataset, we had to increase to 32000 iterations). The BTYD R-package implementation was used for estimating the BG/NBD parameters. Note that the computational costs for deriving P(alive) are negligible for both the ML- and MCMC-based estimation methods.

[15] Note that for $k = 1$, the computational costs for deriving conditional expectations for all ML-based methods are negligible. For $k = 2$, even approximations come at a computation cost, and the CNBD-k models become significantly slower but in practical settings still without problems.

a wider range of practitioners to take advantage of improving the accuracy of their predictions in the presence of regularity.

Against the backdrop of our empirical findings and the diagnostic power of summary statistics, we recommend future model users to routinely quick-check their customer databases for purchase regularity (for example, by using the data-set-level Wheat and Morrison (1990) estimator of regularity). However, the selection of an adequate k for the (M)BG/CNBD-k model should be based on log-likelihood values. Even greater flexibility can be gained by moving to the considerably more computationally demanding Pareto/GGG model (Platzer & Reutterer, 2016). This is particularly advisable if there is probable heterogeneity in regularity across individuals in the customer cohort, paired with higher dropout rates, and enough observations per individual. However, both our simulation and empirical model comparisons reveal that the gain in predictive accuracy of the models developed in this paper comes very close to that of the Pareto/GGG, despite forgoing a customer-specific level of regularity and assuming the degree of regularity to be shared across the whole cohort. Moreover, to fully effectuate its strength in terms of improved predictions, the Pareto/GGG requires sufficient data to update its individual-level regularity estimates. In many empirical settings with large shares of zero-repeat buyers and high dropout rates, this individual-level information is missing, which makes the additional computational effort required by the Pareto/GGG less attractive. From a practical perspective, we conclude that the proposed (M)BG/CNBD-k model families are a good "value-for-effort" compromise, with the chance of increased predictive capacities in the case of regularities and model characteristics that make them scalable and ready-to-implement in data-rich environments.

This research is clearly positioned in the tradition of stochastic models of customer-base analysis and thus shares with them the assumption of stationary marketing activities. It is well documented in the relevant literature that not explicitly modeling non-random marketing actions might result in biased predictions (e.g., Gupta, 1991; Rust et al., 2011; Schweidel & Knox, 2013). For instance, Rust and colleagues demonstrated in an extensive simulation study using, among others, the BG/NBD for predicting future customer profitability that the additional mean absolute percentage error introduced by not modeling marketing interventions is around four to five percentage points (cf. Rust et al., 2011, pp. 289). However, any attempt to incorporate marketing covariates requires access to individual-level full purchase histories or at least a considerably more complex data structure (and a related increase in modeling efforts) than expected from the modest model class we presented in this paper. Moreover, accounting for possible endogeneity of marketing contacts complicates the model structures even more (Rust et al., 2011). In contrast, the event logs of users or summary statistics derived therefrom are easily accessible for any marketing analyst, which makes the models we presented implementable at low cost. In our research, we relaxed another assumption typically made by most stochastic models, namely duration independence or the memoryless assumption of repeat buying. Here, we did this in continuous time, but we expect similar gains in predictive accuracy when accounting for purchase regularity in discrete-time settings. Even though the findings from our simulation studies let us expect the bias introduced by the proposed approximation of the conditional expected number of transactions being very small, it would certainly be interesting to generalize Schmittlein and Morrison's (1983) derivation of an exact closed-form expression for a synchronous counting model combined with latent attrition.

Acknowledgment

The first and the second author gratefully acknowledge financial support from the "ICT of the Future" funding program of the Austrian Federal Ministry for Transport, Innovation and Technology (BMVIT).

Appendix. Derivations

This appendix includes the derivations of the aggregate likelihood functions and key expressions of managerial interest presented in Sections 3.2 and 3.3. for the (M)BG/CNBD-k models.

Appendix A. Aggregate likelihood function

BG/CNBD-k

To derive the likelihood of a randomly chosen customer, we take the expectation of the individual-level results – see Eqs. (1) and (3) – over the gamma mixing distributions for λ (assumption 2) and the beta mixing distribution for p (assumption 4):

$$\begin{aligned} L(k,r,\alpha,a,b|t_1,\ldots,t_x,T) = {} & \delta_{x>0}\,\tilde{t}\int_0^1\int_0^\infty p(1-p)^{x-1}\lambda^{kx}e^{-\lambda t_x}f_\Gamma(\lambda|r,\alpha)f_B(p|a,b)d\lambda dp \\ & + \tilde{t}\int_0^1\int_0^\infty (1-p)^x\lambda^{kx}e^{-\lambda T}\left(\sum_{j=0}^{k-1}\frac{\lambda^j(T-t_x)^j}{j!}\right)f_\Gamma(\lambda|r,\alpha)f_B(p|a,b)d\lambda dp \end{aligned}$$

Model assumption 5 allows us to solve these integrals separately, and we do so by using the notations and results from Hoppe and Wagner (2007):

$$I_\Gamma(i,j,r,\alpha):=\int_0^\infty \lambda^i e^{-\lambda j}f_\Gamma(\lambda|r,\alpha)d\lambda = \frac{\alpha^r}{(\alpha+j)^{r+i}}\frac{\Gamma(r+i)}{\Gamma(r)}$$

$$I_{\mathrm{B}}(i,j,a,b) := \int_0^1 p^i(1-p)^j f_{\mathrm{B}}(p|a,b)dp = \frac{\mathrm{B}(a+i,b+j)}{\mathrm{B}(a,b)}$$

This yields

$$\begin{aligned} &L_{\mathrm{BG/CNBD}-k}(k,r,\alpha,a,b|t_1,\dots,t_x,T) \\ &= \delta_{x>0}\,\tilde{t}\,I_{\mathrm{B}}(1,x-1,a,b)\,I_{\Gamma}(kx,t_x,r,\alpha) + \tilde{t}\,I_{\mathrm{B}}(0,x,a,b)\left(\sum\nolimits_{j=0}^{k-1}\frac{(T-t_x)^j}{j!}I_{\Gamma}(kx+j,T,r,\alpha)\right) \\ &= \delta_{x>0}\,\tilde{t}\frac{\mathrm{B}(a+1,b+x-1)}{\mathrm{B}(a,b)}\frac{\alpha^r}{(\alpha+t_x)^{r+kx}}\frac{\Gamma(r+kx)}{\Gamma(r)} + \tilde{t}\frac{\mathrm{B}(a,b+x)}{\mathrm{B}(a,b)}\left(\sum\nolimits_{j=0}^{k-1}\frac{(T-t_x)^j}{j!}\frac{\alpha^r}{(\alpha+T)^{r+kx+j}}\frac{\Gamma(r+kx+j)}{\Gamma(r)}\right) \end{aligned}$$

MBG/CNBD-k

Applying the same substitutions for the individual-level MBG/CNBD-k likelihood yields

$$\begin{aligned} &L_{\mathrm{MBG/CNBD}-k}(k,r,\alpha,a,b|t_1,\dots,t_x,T) \\ &= \tilde{t}\,I_{\mathrm{B}}(1,x,a,b)\,I_{\Gamma}(kx,t_x,r,\alpha) + \tilde{t}\,I_{\mathrm{B}}(0,x+1,a,b)\left(\sum\nolimits_{j=0}^{k-1}\frac{(T-t_x)^j}{j!}I_{\Gamma}(kx+j,T,r,\alpha)\right) \\ &= \tilde{t}\frac{\mathrm{B}(a+1,b+x)}{\mathrm{B}(a,b)}\frac{\alpha^r}{(\alpha+t_x)^{r+kx}}\frac{\Gamma(r+kx)}{\Gamma(r)} + \tilde{t}\frac{\mathrm{B}(a,b+x+1)}{\mathrm{B}(a,b)}\left(\sum\nolimits_{j=0}^{k-1}\frac{(T-t_x)^j}{j!}\frac{\alpha^r}{(\alpha+T)^{r+kx+j}}\frac{\Gamma(r+kx+j)}{\Gamma(r)}\right) \end{aligned}$$

Appendix B. Probability of being alive

BG/CNBD-k

Deriving the probability of being alive is straightforward and follows the main argument from Fader et al.'s (2008) research note on deriving $P(\text{alive})$ for the BG/NBD model. The summands of the aggregate-level likelihood already reflect the two cases, being alive at T and having dropped out at t_x. Thus, by simply applying Bayes' Theorem we can express the probability of being alive as

$$P(\text{alive}) = \frac{L(\dots|\text{alive})}{L(\dots)} = \frac{L(\dots|\text{alive})}{L(\dots|\mathit{alive}) + L(\dots|\text{dead})} = \left(1 + \frac{L(\dots|\text{dead})}{L(\dots|\text{alive})}\right)^{-1}$$

By substituting

$$L(k,r,\alpha,a,b|\text{dead}) = \delta_{x>0}\,\tilde{t}\frac{\mathrm{B}(a+1,b+x-1)}{\mathrm{B}(a,b)}\frac{\alpha^r}{(\alpha+t_x)^{r+kx}}\frac{\Gamma(r+kx)}{\Gamma(r)}$$

$$L(k,r,\alpha,a,b|\text{alive}) = \tilde{t}\frac{\mathrm{B}(a,b+x)}{\mathrm{B}(a,b)}\left(\sum\nolimits_{j=0}^{k-1}\frac{(T-t_x)^j}{j!}\frac{\alpha^r}{(\alpha+T)^{r+kx+j}}\frac{\Gamma(r+kx+j)}{\Gamma(r)}\right)$$

and using the relationship

$$B(a+1,b) = \frac{a}{a+b}B(a,b)$$

we can reduce the fraction and simplify to

$$P_{\mathrm{BG/CNBD}-k}(\text{alive at }T|t_1,\dots,t_x,T,k,r,\alpha,a,b) = \left(1 + \delta_{x>0}\frac{a}{b+x-1}\left(\frac{\alpha+T}{\alpha+t_x}\right)^{r+kx} \Big/ \sum\nolimits_{j=0}^{k-1}\frac{(T-t_x)^j}{j!(\alpha+T)^j}\frac{\Gamma(r+kx+j)}{\Gamma(r)}\right)^{-1}$$

MBG/CNBD-k

In analogous manner, we can simplify the fraction of the likelihood summands for the MBG/CNBD-k model to

$$P_{\text{MBG/CNBD}-k}(\text{alive at } T|t_1,\ldots,t_x,T,k,r,\alpha,a,b) = \left(1+\frac{a}{b+x}\left(\frac{\alpha+T}{\alpha+t_x}\right)^{r+kx} / \sum_{j=0}^{k-1}\frac{(T-t_x)^j}{j!(\alpha+T)^j}\frac{\Gamma(r+kx+j)}{\Gamma(r)}\right)^{-1}$$

Appendix C. Unconditional probability distribution of purchase frequencies

BG/CNBD-k

The probability of a randomly chosen customer encountering x transactions within time T is derived following Fader et al. (2005a):

$$P(X(t)=x) = P(\text{alive after x}-\text{th purchase})\cdot P\left(t_x \le t \text{ and } t_{x+1} > t\right) + \delta_{x>0}\, P(\text{dropout at x}-\text{th purchase})\cdot P(t_x \le t)$$

$$P(X(t)=x|k,\lambda,p) = (1-p)^x \sum_{j=kx}^{kx+k-1} P_{\text{Poisson}}(X(t)=j|\lambda,p) + \delta_{x>0}\, p\,(1-p)^{x-1}\left(1-\sum_{j=0}^{kx+k-1} P_{\text{Poisson}}(X(t)=j|\lambda,p)\right)$$

Substituting the Poisson pdf and integrating over the mixing distributions yields

$$\begin{aligned}P_{\text{BG/CNBD}-k}(X(t)=x|k,r,\alpha,a,b) = {} & \frac{\mathrm{B}(a,b+x)}{\mathrm{B}(a,b)}\sum_{j=kx}^{kx+k-1}\frac{t^j}{j!}\frac{\alpha^r}{(\alpha+t)^{r+j}}\frac{\Gamma(r+j)}{\Gamma(r)}\\ & + \delta_{x>0}\,\frac{\mathrm{B}(a+1,b+x-1)}{\mathrm{B}(a,b)}\left(1-\sum_{j=0}^{kx+k-1}\frac{t^j}{j!}\frac{\alpha^r}{(\alpha+t)^{r+j}}\frac{\Gamma(r+j)}{\Gamma(r)}\right)\end{aligned}$$

Note, that the summands are the pdf of the Negative Binomial distribution (NBD):

$$P_{\text{NBD}}(X(t)=j|r,\alpha) = \frac{t^j}{j!}\frac{\alpha^r}{(\alpha+t)^{r+j}}\frac{\Gamma(r+j)}{\Gamma(r)}$$

MBG/CNBD-k

Again, the derivations only differ in terms of one additional dropout opportunity at time $t_{0:}$

$$\begin{aligned}P_{\text{MBG/CNBD}-k}(X(t)=x|k,r,\alpha,a,b) = {} & \frac{\mathrm{B}(a,b+x+1)}{\mathrm{B}(a,b)}\sum_{j=kx}^{kx+k-1}\frac{t^j}{j!}\frac{\alpha^r}{(\alpha+t)^{r+j}}\frac{\Gamma(r+j)}{\Gamma(r)}\\ & + \frac{\mathrm{B}(a+1,b+x)}{\mathrm{B}(a,b)}\left(1-\sum_{j=0}^{kx+k-1}\frac{t^j}{j!}\frac{\alpha^r}{(\alpha+t)^{r+j}}\frac{\Gamma(r+j)}{\Gamma(r)}\right)\end{aligned}$$

Appendix D. Unconditional expectation for condensed Poisson

The expected number of transactions for an active customer with exponentially distributed interevent times is known to be $E(X(t)|\lambda) = \lambda t$. The asynchronous counting process for Erlang-2 waiting times has an expectation of $E(X(t)|\lambda) = \lambda t/2$ (Chatfield & Goodhardt, 1973, p. 829). Similarly, we will now prove that the generalization for Erlang-k $E(X(t)|k,\lambda) = \lambda t/k$ also holds true. Recall that asynchronous counting for Erlang-k can also be seen as a censored counting of a Poisson process, where every kth event is being counted. As we begin the counting independent of a particular initial event, the recording of r censored events can either arise from recording rk, $rk+1$, $rk-1$, …, $rk+k-1$ *or* $rk-k+1$ uncensored events. Alternatively, from the opposite perspective, $rk+j$ uncensored events result with probability $\frac{k-j}{k}$ in r and with probability $\frac{j}{k}$ in $r+1$ censored events to be counted (for

$0 \le j \le k-1$). Using the compact notations $P(r)$ for $P(X(t)=r|k,\lambda)$ and $P_P(r)$ for $P_{Poisson}(X(t)=r|\lambda)$, we can therefore write the expectation as

$$\begin{aligned} E(X(t)|k,\lambda) &= \sum_{r=1}^{\infty} r\,P(r) = \sum_{r=1}^{\infty} r\left(\sum_{j=0}^{k-1}\frac{j}{k}P_P(rk-k+j) + \sum_{j=0}^{k-1}\frac{k-j}{k}P_P(rk+j)\right) \\ &= \frac{1}{k}\sum_{r=1}^{\infty} rk\,P_P(rk) + \sum_{j=1}^{k-1}\underbrace{\left(\frac{j}{k^2}\sum\nolimits_{r=1}^{\infty} rk\,P_P(rk-k+j) + \frac{k-j}{k^2}\sum\nolimits_{r=1}^{\infty} rk\,P_P(rk+j)\right)}_{=:T_j} \end{aligned}$$

T_j can be reduced to

$$\begin{aligned} T_j &= \frac{j}{k^2}\left(\sum(rk-k+j)P_P(rk-k+j) + \sum(k-j)P_P(rk-k+j)\right) + \frac{k-j}{k^2}\left(\sum(rk+j)P_P(rk+j) - \sum jP_P(rk+j)\right) \\ &= \frac{j}{k^2}\left(\sum(rk-k+j)P_P(rk-k+j) + \sum(k-j)P_P(rk-k+j)\right) \\ &\quad + \frac{k-j}{k^2}\left(\sum(rk-k+j)P_P(rk-k+j) - jP_P(j) - \sum jP_P(rk-k+j) + jP_P(j)\right) \\ &= \frac{1}{k}\sum(rk-k+j)P_P(rk-k+j) \end{aligned}$$

and we receive our previously stated result for the unconditional expected number for asynchronous counting:

$$E(X(t)|k,\lambda) = \frac{1}{k}\sum\nolimits_{r=1}^{\infty} rk\,P_P(rk) + \sum\nolimits_{j=1}^{k-1}\frac{1}{k}\sum\nolimits_{r=1}^{\infty}(rk-k+j)P_P(rk-k+j) = \frac{1}{k}\sum\nolimits_{r=1}^{\infty} r\,P_P(r) = \frac{\lambda t}{k}$$

References

Allenby, G. M., Leone, R. P., & Jen, L. (1999). A dynamic model of purchase timing with application to direct marketing. *Journal of the American Statistical Association*, *94*(446), 365–374.
Batislam, E. P., Denizel, M., & Filiztekin, A. (2007). Empirical validation and comparison of models for customer base analysis. *International Journal of Research in Marketing*, *24*(3), 201–209.
Bemmaor, A. C., & Glady, N. (2012). Modeling purchasing behavior with sudden "death": A flexible customer lifetime model. *Management Science*, *58*(5), 1012–1021.
Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Wadsworth.
Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, *16*(5), 1190–1208.
Chatfield, C., & Goodhardt, G. J. (1973). A consumer purchasing model with Erlang inter-purchase times. *Journal of the American Statistical Association*, *68*(344), 828–835.
Driscoll, M. (2017). *Customer Lifetime Value. The Robin Report* (2nd August 2017) available viahttp://www.therobinreport.com/customer-lifetime-value/.
Fader, P. S., & Hardie, B. G. (2001). Forecasting repeat sales at CDNOW: A case study. *Interfaces*, 31(3_supplement), S94-S107.
Fader, P. S., & Hardie, B. G. (2009). Probability models for customer-base analysis. *Journal of Interactive Marketing*, *23*(1), 61–69.
Fader, P. S., Hardie, B. G., & Huang, C. Y. (2004). A dynamic changepoint model for new product sales forecasting. *Marketing Science*, *23*(1), 50–65.
Fader, P. S., Hardie, B. G., & Lee, K. L. (2005a). "Counting your customers" the easy way: An alternative to the Pareto/NBD model. *Marketing Science*, *24*(2), 275–284.
Fader, P. S., Hardie, B. G., & Lee, K. L. (2005b). RFM and CLV: Using iso-value curves for customer base analysis. *Journal of Marketing Research*, *42*(4), 415–430.
Fader, P. S., Hardie, B.G., & Lee, K. L. (2008). Computing P (alive) using the BG/NBD model. Research Note available via http://www.brucehardie.com/notes/021.
Fader, P. S., Hardie, B. G., & Shang, J. (2010). Customer-base analysis in a discrete-time noncontractual setting. *Marketing Science*, *29*(6), 1086–1108.
Gupta, S. (1988). Impact of sales promotions on when, what, and how much to buy. *Journal of Marketing Research*, 342–355.
Gupta, S. (1991). Stochastic models of interpurchase time with time-dependent covariates. *Journal of Marketing Research*, *28*(1), 1–15.
Gupta, S., Hanssens, D., Hardie, B., Kahn, W., Kumar, V., Lin, N., & Sriram, S. (2006). Modeling customer lifetime value. *Journal of Service Research*, *9*(2), 139–155.
Herniter, J. (1971). A probabilistic market model of purchase timing and brand selection. *Management Science*, *18*(4, part-II), 102–113.
Hoppe, D., & Wagner, U. (2007). Customer base analysis: The case for a central variant of the Betageometric/NBD model. *Marketing ZFP*, *29*(JRM 2), 75–90.
Jerath, K., Fader, P. S., & Hardie, B. G. (2011). New perspectives on customer "death" using a generalization of the Pareto/NBD model. *Marketing Science*, *30*(5), 866–880.
Kahn, B. E. (1987). A theoretical model of interpurchase times. *Applied Stochastic Models in Business and Industry*, *3*(2), 93–109.
Kim, B. D., & Park, K. (1997). Studying patterns of consumer's grocery shopping trip. *Journal of Retailing*, *73*(4), 501–517.
Ma, S. H., & Liu, J. L. (2007, August). The MCMC approach for solving the Pareto/NBD model and possible extensions. In Natural Computation, 2007. ICNC 2007. Third International Conference on (Vol. 2, pp. 505-512). IEEE.
Malthouse, E. C. (2009). The results from the lifetime value and customer equity modeling competition. *Journal of Interactive Marketing*, *23*(3), 272–275.
McShane, B., Adrian, M., Bradlow, E. T., & Fader, P. S. (2008). Count models based on Weibull interarrival times. *Journal of Business & Economic Statistics*, *26*(3), 369–378.
Morrison, D. G., & Schmittlein, D. C. (1981). Predicting future random events based on past performance. *Management Science*, *27*(9), 1006–1023.
Morrison, D. G., & Schmittlein, D. C. (1988). Generalizing the NBD model for customer purchases: What are the implications and is it worth the effort? *Journal of Business & Economic Statistics*, *6*(2), 145–159.
Mzoughia, M. B., Borle, S., & Limam, M. (2018). A MCMC approach for modeling customer lifetime behavior using the COM-Poisson distribution. *Applied Stochastic Models in Business and Industry*, *34*(2), 113–127.
Mzoughia, M. B., & Limam, M. (2014). An improved BG/NBD approach for modeling purchasing behavior using COM-Poisson distribution. *International Journal of Modeling and Optimization*, *4*(2), 141–145.
Park, C. H. (2017). Online purchase paths and conversion dynamics across multiple websites. *Journal of Retailing*, *93*(3), 253–265.

Park, C. H., & Park, Y. H. (2016). Investigating purchase conversion by uncovering online visit patterns. *Marketing Science*, *35*(6), 894–914.
Platzer, M., & Reutterer, T. (2016). Ticking away the moments: Timing regularity helps to better predict customer activity. *Marketing Science*, *35*(5), 779–799.
Reinartz, W. J., & Kumar, V. (2000). On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing. *Journal of Marketing*, *64*(4), 17–35.
Rust, R. T., Kumar, V., & Venkatesan, R. (2011). Will the frog change into a prince? Predicting future customer profitability. *International Journal of Research in Marketing*, *28*(4), 281–294.
Schmittlein, D. C., & Morrison, D. G. (1983). Prediction of future random events with the condensed negative binomial distribution. *Journal of the American Statistical Association*, *78*(382), 449–456.
Schmittlein, D.C., Morrison, D. G., & Colombo, R. (1987). Counting your customers: Who-are they and what will they do next? *Management Science*, *33*(1), 1–24.
Schwartz, E. M., Bradlow, E. T., & Fader, P. S. (2014). Model selection using database characteristics: Developing a classification tree for longitudinal incidence data. *Marketing Science*, *33*(2), 188–205.
Schweidel, D. A., & Fader, P. S. (2009). Dynamic changepoints revisited: An evolving process model of new product sales. *International Journal of Research in Marketing*, *26*(2), 119–124.
Schweidel, D. A., & Knox, G. (2013). Incorporating direct marketing activity into latent attrition models. *Marketing Science*, *32*(3), 471–487.
Schweidel, D. A., & Moe, W. W. (2016). Binge watching and advertising. *Journal of Marketing*, *80*(5), 1–19.
Therneau, T., Atkinson, B., & Ripley, B. (2019). *rpart: Recursive Partitioning and Regression Trees.* Available via https://CRAN.R-project.org/package=rpart4. (pp. 1–15). R package version, 1–15.
Wagner, U., & Hoppe, D. (2008). Erratum on the MBG/NBD model. *International Journal of Research in Marketing*, *25*(3), 225–226.
Wheat, R. D., & Morrison, D. G. (1990). Estimating purchase regularity with two interpurchase times. *Journal of Marketing Research*, *27*(1), 87–93.
Winkelmann, R. (1995). Duration dependence and dispersion in count-data models. *Journal of Business & Economic Statistics*, *13*(4), 467–474.
Wu, C., & Chen, H. L. (2000a). A consumer purchasing model with learning and departure behaviour. *Journal of the Operational Research Society*, 583–591.
Wu, C., & Chen, H. L. (2000b). Counting your customers: Compounding customer's in-store decisions, interpurchase time and repurchasing behavior. *European Journal of Operational Research*, *127*(1), 109–119.
Zhang, Y., Bradlow, E. T., & Small, D. S. (2013). New measures of clumpiness for incidence data. *Journal of Applied Statistics*, *40*(11), 2533–2548.
Zhang, Y., Bradlow, E. T., & Small, D. S. (2015). Predicting customer value using clumpiness: From RFM to RFMC. *Marketing Science*, *34*(2), 195–208.