



## Marketing Science

Publication details, including instructions for authors and subscription information:  
<http://pubsonline.informs.org>

### Ticking Away the Moments: Timing Regularity Helps to Better Predict Customer Activity

Michael Platzer, Thomas Reutterer

To cite this article:

Michael Platzer, Thomas Reutterer (2016) Ticking Away the Moments: Timing Regularity Helps to Better Predict Customer Activity. Marketing Science 35(5):779-799. <https://doi.org/10.1287/mksc.2015.0963>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact [permissions@informs.org](mailto:permissions@informs.org).

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2016, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

# Ticking Away the Moments: Timing Regularity Helps to Better Predict Customer Activity

Michael Platzer, Thomas Reutterer

Department of Marketing, WU Vienna University of Economics and Business, A-1020 Vienna, Austria  
{[michael.platzer@gmail.com](mailto:michael.platzer@gmail.com), [thomas.reutterer@wu.ac.at](mailto:thomas.reutterer@wu.ac.at)}

Accurate predictions of a customer's activity status and future purchase propensities are crucial for managing customer relationships. This article extends the recency–frequency paradigm of customer-base analysis by integrating regularity in interpurchase timing in a modeling framework. By definition, regularity implies less variation in timing patterns and thus better predictability. Whereas most stochastic customer behavior models assume a Poisson process of “random” purchase occurrence, allowing for regularity in the purchase timings is beneficial in noncontractual settings because it improves inferences about customers' latent activity status. This especially applies to those valuable customers who were previously very frequently active but have recently exhibited a longer purchase hiatus. A newly developed generalization of the well-known Pareto/NBD model accounts for varying degrees of regularity across customers by replacing the NBD component with a mixture of gamma distributions (labeled Pareto/GGG). The authors demonstrate the impact of incorporating regularity on forecasting accuracy using an extensive simulation study and a range of empirical applications. Even for mildly regular timing patterns, it is possible to improve customer-level predictions; the stronger the regularity, the greater the gain. Furthermore, the cost in terms of data requirements is marginal because only one additional summary statistic, in addition to recency and frequency, is needed that captures historical transaction timing.

Data, as supplemental material, are available at <http://dx.doi.org/10.1287/mksc.2015.0963>.

**Keywords:** customer-base analysis; customer lifetime value; purchase regularity; stochastic prediction models; noncontractual settings; Pareto/NBD

**History:** Received: August 12, 2013; accepted: August 17, 2015; Preyas Desai served as the editor-in-chief and Peter Fader served as associate editor for this article. Published online in *Articles in Advance* May 3, 2016.

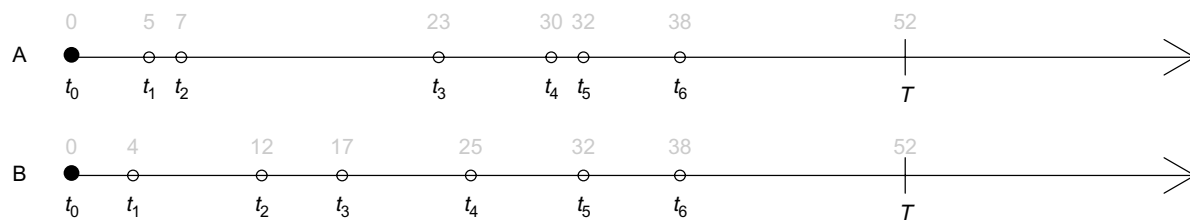
## 1. Introduction

The song “Time” by world-famous rock band Pink Floyd is about how a lifetime can rush by and that many people do not realize this until it is too late. This anecdotal evidence by Pink Floyd's songwriter Roger Waters also mirrors some facets of enduring, but noncontractual, customer–firm relationships. In such settings, managers are often left with an uncomfortable degree of ambiguity concerning how to react to the moments that are ticking away with customers who were previously active but have recently grown “silent” in interacting with the focal firm. This paper proposes to consider past *timing patterns* when interpreting a customer's most recent activity hiatus. By introducing a new probabilistic model for customer-base analysis, the Pareto/GGG, we will show that regularity in interevent timings helps to better predict customer activity.

In recent years, customer-base analysis has become increasingly popular among marketing analysts and data scientists (Winer 2001, Fader and Hardie 2009). Triggered by the availability of vast amounts of customer-level transaction data, combined with the facilitated access to and greater usability of sophisticated models, this popularity is also paralleled by the growing managerial interest in residual customer

lifetime value (CLV) and its subcomponents as key metrics for managing customer-centric organizations (Shah et al. 2006, Tirenni et al. 2007, Kumar et al. 2008, Kumar 2008, Fader 2013). One of the main challenges in such analyses is accurately predicting future purchase behavior when customer–firm relationships are of a *noncontractual* nature (Reinartz and Kumar 2000, Gupta et al. 2006). First, in such a setting, a customer's current status at time  $T$  is not directly observable by the organization, but must be inferred indirectly from past activity. Second, the available historical purchase data is right censored at time  $T$ , such that the full lifetime of a customer cohort has yet to be observed. Third, the amount of customer-level data tends to vary significantly. In general, despite the richness of the data in the aggregate, only a few transactions are observed for most customers, and hence to extract the most information from the available data, marketing analysts need to adaptively pool information across customers.

Figure 1 illustrates the transaction records of two hypothetical customers in a noncontractual setting; the solid black dots indicate the first purchases, and the circles represent repeat purchases. Both cases exhibit identical values of the two widely used statistics to

**Figure 1** Same Recency, Same Frequency, yet Customer A Is More Likely to Be Alive than Customer B

summarize observed transaction histories, namely, *recency* ( $R$ , indicating the date of the last transaction; here,  $t_6 = 38$ ) and *frequency* ( $F$ , or the total number of transactions; here,  $x = 6$ ). The key question for managers is to determine which customer is more valuable to the company at the current time  $T = 52$  and thereafter. Applying a purely recency-based heuristic, such as the “hiatus heuristic” used by Wübben and von Wangenheim (2008), would leave managers undecided. The same applies to simple RF-based customer scoring models (Malthouse 2001, Malthouse and Blattberg 2005), which were originally imported from direct marketing and remain popular among many practitioners for valuing their customers. A series of increasingly sophisticated probabilistic models link simple RF summary statistics with a theoretically well-grounded behavioral story of customers’ repurchase behavior and explicitly integrate latent customer attrition into the modeling framework (Fader and Hardie 2009). These so-called “buy-till-you-die” (BTYD) models represent the well-established standard approach used by data scientists and analysts to predict crucial CLV components (such as the expected number of future purchases) and have been applied in a wide variety of industries (Wübben and von Wangenheim 2008).<sup>1</sup>

In a standard BTYD model, however, customers A and B would also be evaluated equally (same  $R$  and  $F$ ), which is a direct consequence of the model’s assumptions regarding the purchase process of active customers. For example, the most widely recognized benchmark BTYD model, the Pareto/NBD (Schmittlein et al. 1987), assumes a Poisson purchase process and an exponentially distributed lifetime. The two corresponding customer-level parameters can vary across customers, following independent gamma distributions. Because the gamma-exponential mixture results in a Pareto distribution, whereas the gamma-Poisson mixture leads to a negative binomial distribution, the model is referred to as Pareto/NBD. Among the many variations and extensions of the Pareto/NBD model (Fader and Hardie 2009), most retain a Poisson purchase

process, in which the time between two purchases is exponentially distributed. This implies that the most likely time for a repurchase is immediately after a purchase (mode zero). Furthermore, because of the memoryless property of the exponential distribution, the time elapsed since the last purchase does not influence the timing of the next purchase (Chatfield and Goodhardt 1973).

Because of these properties, NBD-based models interpret the purchase hiatus at the end of the observation period equally for both customers in Figure 1. However, their observed intertransaction timing patterns tell a different story: Customer B shows a regular repurchase pattern, with narrowly distributed intertransaction times (ITTs). The long waiting time since the last transaction thus differs from that customer’s individual norm, suggesting that customer B may have actually defected. By contrast, the recent purchase hiatus exhibited by customer A is observed before  $(t_3 - t_2)$ , and hence it does not provide a similarly clear indication of whether this person will be active in the future. A model that adequately accounts for these differences in timing patterns would thus assign a significantly lower probability of future activities to the regular but “overdue” customer relative to the customer purchasing more irregularly. Therefore, these two customers also deserve to be treated differently in terms of allocating marketing resources and tailoring marketing campaigns to them. Efforts need to be undertaken to win back customer B, whereas customer A could be targeted with up- and cross-selling opportunities, because the next purchase event is anticipated to take place soon.

Recently, in a similar vein, but with opposite signs, Zhang et al. (2015) introduced the concept of *clumpiness* in the marketing literature and proposed extending the  $R$ ,  $F$ , monetary value (RFM)-based framework for CLV predictions by including a metric-based approach that captures variations in timing patterns across customers. By definition, their proposed clumpiness metric  $C$  takes its minimum value with equally spaced events; hence, clumpiness can be understood as the opposite of regularity. Such non-Poisson-like, “clumpy” patterns of rapidly occurring events separated by longer periods of inactivity can occur, for example, in digital media consumption and website visits (e.g., YouTube, Amazon, eBay, Hulu), where many consumers tend to “binge.”

<sup>1</sup> The popularity of BTYD models also has benefited greatly from Excel-based implementations (Fader et al. 2005a) and the more recent availability of open-source implementations in more sophisticated programming environments, such as the R package BTYD (Dziurzynski et al. 2014).

Barabasi (2005) lists further examples of certain human activities that exhibit such “bursts” of activity, followed by longer periods of inactivity, and shows that such patterns can arise as a consequence of a priority-driven decision process. However, there is also evidence that more regular timing patterns are prevalent in other contexts (Chatfield and Goodhardt 1973, Gupta 1991, Wu and Chen 2000, Bardhan et al. 2015), such as purchases of packaged consumer goods that are frequently consumed (e.g., food, beverages, detergents, toiletries). The consumption patterns for these products are often regular per se and thus trigger regular purchases. Furthermore, if many acts of consumption are required to initiate the need for repurchase (e.g., breakfast cereals, ground coffee, toothpaste, toilet paper), then even if consumption timing is random, purchase timing still will appear regular. Intuitively, this regularity results when variations in interconsumption times cancel one another out when summed together.<sup>2</sup>

In this paper, we respond to the call by Zhang et al. (2015) to develop a model-based approach that accommodates intertransaction timing patterns in predicting future purchase activities. Our proposed model builds on a probabilistic BTYD framework and leverages information on timing patterns to make improved inferences regarding customers’ activity status. We develop a Pareto/NBD variant that replaces the NBD repeat-buying component with a mixture of gamma distributions (Pareto/GGG) to allow for a varying degree of regularity across customers. We thereby abandon the historical focus on the *count process*, but fully emphasize accurately capturing the *timing process*. Although the model captures a wide variety of timing patterns, it does not induce any substantial additional costs in terms of data requirements. Other than the usual RF statistics, it requires only a single summary statistic for the historical ITTs, which can be provided easily while updating the RF variables.

Before we present the formal properties of the model in §3, we review prior contributions that suggest relaxing the restrictive intertransaction timing assumptions underlying conventional BTYD models and thereby draw conclusions for developing our model variant. We show that the proposed model generalizes the Pareto/NBD by accommodating regular timing patterns, but also nests cases of random (i.e., exponentially distributed) and clumpy patterns. In §4, we explore the improved predictive performance of the Pareto/GGG using a like-for-like comparison with the Pareto/NBD across a broad range of simulated parameter settings, then investigate for which customer segments we should expect the greatest improvement in holdout

predictions. Next, we empirically validate the forecasting performance of our model using multiple data sets. By accounting for regularity, our model outperforms both the Pareto/NBD and the heuristic benchmark suggested by Wübben and von Wangenheim (2008) in terms of out-of-sample, individual-level forecasting accuracy for future customer-level transactions. This is already the case for relatively mild intertransaction timing regularities, whereas the models perform on par for data sets without regularities. When ITTs feature regularity, the greatest improvements accrue in the important, high-frequency, low-recency customer segment. In such a situation, a standard NBD-type model results in overly optimistic predictions and erroneous recommendations with respect to customer prioritization. In §5, we contrast the properties of our model-based approach with the clumpiness metric  $C$  developed by Zhang et al. (2015) in greater detail. Finally, we discuss the merits and limitations of the proposed model for researchers and practitioners and outline some suggestions for further research.

## 2. Random, Regular, and Clumpy Interpurchase Timing

Since its introduction, the Pareto/NBD model has been extended in various ways, mostly by modifying the dropout process. A particularly noteworthy variation resulted in the betageometric/NBD model (BG/NBD) by Fader et al. (2005a), which adjusts the dropout story by restricting defection to repurchase incidents. The BG/NBD approach offers data-fitting capabilities similar to those of the Pareto/NBD model, but it is mathematically and computationally less demanding, which has helped disseminate this model class in real-world settings. In turn, Batislam et al. (2007) and Hoppe and Wagner (2007) each modified this variant by allowing for an additional dropout opportunity immediately after the initial purchase (MBG/NBD and CBG/NBD, respectively). In the model developed by Jerath et al. (2011), *periodic death opportunities* serve to decouple the discrete dropout opportunities from the purchase process. Bemmaor and Glady (2012) reintroduces the assumption of a continuously distributed dropout process, but allows for a nonconstant hazard rate (gamma/Gompertz/NBD). Furthermore, other Pareto/NBD variants have succeeded in incorporating time-invariant covariates (Fader and Hardie 2007, Abe 2009). However, all of these models retain the assumptions of a Poisson purchase process (i.e., NBD). Considering that the dropout process is latent, whereas the purchase process is directly observable and offers more customer-level information, it seems appropriate to focus on the purchase process and model a more flexible distribution to adapt to a wider range of real-world timing patterns. This adaptation appears particularly necessary against the backdrop of our research

<sup>2</sup> Specifically, the sum of  $k$  independent exponential variables follows an Erlang- $k$  distribution, and its coefficient of variation decreases with increasing  $k$ :  $CV = 1/\sqrt{k}$ .



motivation, namely, to gain a better understanding of the purchase process and thereby to also improve inferences concerning the nonobservable activity status.

The idea of accounting for deviations from Poisson-like purchasing in repeat-buying models is not new in the marketing literature. Herniter (1971) was among the first to propose modeling ITTs using the Erlang- $k$  family of distributions, which allows for various degrees of regularity in timing patterns (i.e., higher  $k$ , stronger regularity), but also contains the exponential distribution as a special case ( $k = 1$ ). In addition, Chatfield and Goodhardt (1973) combined the Erlang- $k$  with a gamma distribution to reflect the variation in the purchase rate across customers and termed the resulting model a *condensed negative binomial distribution* (CNBD). Both contributions also offered empirical support for regular purchase patterns of  $k = 2$  for products such as aluminum foil, detergents, razor blades, and soaps. However, in their concluding remarks, Chatfield and Goodhardt (1973, p. 834) worried that “the CNBD formulas are so much more complex that it is doubtful if the small improvements in fit that seem possible justify the extra effort.” Furthermore, these authors focused on stationary settings in a repeat-buying context, whereas we attempt to leverage a better understanding of the timing process to improve inferences concerning the latent activity status. Further research on the characteristics of the CNBD appear in the papers by Schmittlein and Morrison (1983) and Morrison and Schmittlein (1988). Gupta (1991) presents a framework capable of incorporating time-dependent covariates for estimating NBD as well as CNBD models; the empirical evidence in that article also suggests regular purchase patterns for the coffee category. Within this framework, Fader et al. (2004) build a dynamic changepoint model, in which changes in purchase frequency predict new product sales along the product cycle. In addition, Schweidel and Fader (2009) allow for a transition from an exponentially distributed to a more regular Erlang-2 timing pattern over a customer’s life cycle. Finally, Wu and Chen (2000) combine the CNBD model with a nonstationary repeat-buying process and observe strong regularities in the tea category, with an estimate of  $k = 5$ .

It is possible to achieve greater flexibility for modeling intertransaction timing regularity by moving from Erlang- $k$  to a gamma distribution, which contains the former as a special case, although its shape parameter is no longer restricted to integer values. The shape parameter can be estimated at the customer level by calculating the coefficient of variation (CV) or maximum likelihood and is then aggregated across customers (Herniter 1971, Dunn et al. 1983, Wu and Chen 2000). However, this approach requires sufficient transactions made by a customer (usually 10 or more) and thus suffers in settings marked by low frequency.

Wheat and Morrison (1990) propose an alternative estimation method that requires only two observed ITTs ( $\Delta t_1, \Delta t_2$ ) per customer by assuming a common shape parameter  $k$  across all customers. The estimate for this shape parameter is given by

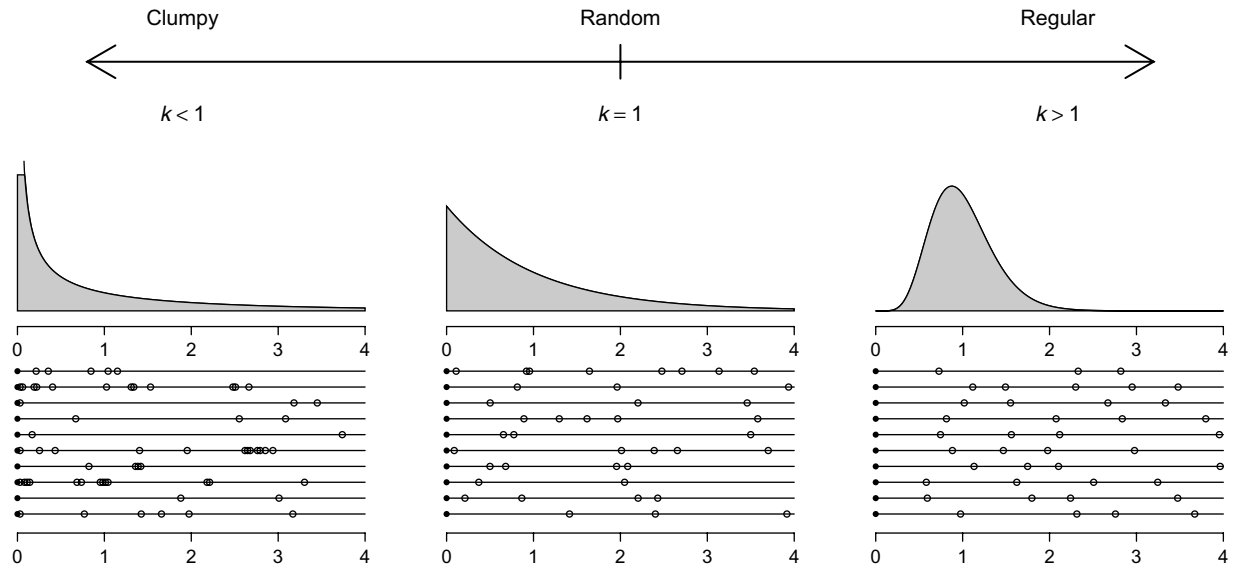
$$\hat{k}_{\text{wheat}} = \frac{1 - 4 \cdot \text{var}(M)}{8 \cdot \text{var}(M)} \quad \text{with } M := \frac{\Delta t_1}{\Delta t_1 + \Delta t_2}, \quad (1)$$

which serves as an easy-to-compute data-set-level summary statistic and can be used as a quick diagnostic check for regularity within a customer base before running a more complex modeling approach, such as the Pareto/GGG presented in the next section. In addition to the above work, Allenby et al. (1999) present an even more flexible model that assumes that the intertransaction timing patterns follow a generalized gamma distribution, which contains the gamma, Weibull, lognormal, Erlang, and exponential distributions as special cases, and hence it accommodates a wide variety of timing patterns, including regular purchases. However, both shape parameters remain constant across customers in their model; thus, they only allow for a homogeneous regularity parameter in the customer base.

More recently, Zhang et al. (2013) introduce an entire class of  $C$  measures to calculate the degree of “clumpiness” at an individual level in time-discrete settings. The design of these measures has been guided by four requirements concerning their behavior. Among others, they are expected to take their maxima when events are tightly clustered and their minima in the case of constant ITTs. Thus, they effectively capture the level of nonrandomness in event timings and also measure regularity, just with opposite signs. Nevertheless, as we will show in §5, to become reliable, these metrics also require a high number of observations at the individual level and therefore need to be used with care in low-frequency or high-churn scenarios. The generalization of the Pareto/NBD presented in the next section, the Pareto/GGG, is able to address both situations well, because it adaptively pools available information across customers and explicitly allows for churn to take place.

Before introducing this approach, we depict the range of timing patterns induced by assuming gamma-distributed ITTs. Figure 2 displays the probability density and multiple sampled timing patterns for three different values of the shape parameter ( $k \in \{0.3, 1, 8\}$ ). To facilitate direct comparisons across settings, we also use  $k$  as a rate parameter, such that all result in the same expected ITT of one time unit. Obviously, for higher values of  $k$ , the density exhibits a narrower shape, resulting in less variance in ITTs and thus more regular patterns. For smaller values of  $k$ , we instead detect tight clusters of transactions along the timeline. As this brief illustration shows, the gamma distribution

**Figure 2** Distribution and Sampled Timing Patterns for Gamma( $k, k$ )



can capture regular ( $k > 1$ ), random ( $k = 1$ ), and clumpy ( $k < 1$ ) patterns; it generalizes the exponential and Erlang- $k$  family of distributions and thus represents a good candidate for a flexible model of timing patterns in noncontractual settings with continuous timing.

### 3. Model Development

Because the Pareto/NBD model remains the workhorse among BTYD models operating in continuous time (Wübben and von Wangenheim 2008), we selected it as the base for developing a more general model that can account for various timing patterns. Specifically, we replace the exponential with a gamma distribution to model a customer's ITTs and let the shape parameter of that gamma distribution vary across customers. In so doing, we not only allow for heterogeneity in frequency and dropout but also in terms of regularity. Note that other NBD-based models could be generalized in a similar way. Using a hierarchical Bayes setup, the model adaptively pools (the commonly sparse) individual-level information, and the degree of pooling is driven by the data (Rossi and Allenby 2003). Thus, a prior belief about a customer's regularity gets updated in a Bayesian manner, according to the available customer-level information about variance in ITTs. The more transactions that are available for a customer, the better we can draw inferences on that customer's regularity.

#### 3.1. Model Assumptions

**3.1.1. Transaction Process.** While the customer remains alive, her ITTs,  $\Delta t_j := t_j - t_{j-1}$ , follow a gamma distribution, with shape parameter  $k$  and rate parameter  $k\lambda$ :

$$\Delta t_j \sim \text{Gamma}(k, k\lambda). \quad (2)$$

The mean of the gamma distribution is shape/rate, and the CV is  $1/\sqrt{\text{shape}}$ . The chosen parameterization therefore results in the same mean ITT as the Pareto/NBD,  $1/\lambda$ , which allows for a direct comparison of the parameter estimates of  $\lambda$  between the two models. Here,  $\lambda$  determines the frequency, and the shape parameter  $k$  determines the regularity of the transaction timings.

**3.1.2. Dropout Process.** A customer remains alive for an exponentially distributed lifetime  $\tau$

$$\tau \sim \text{Exponential}(\mu). \quad (3)$$

**3.1.3. Heterogeneity Across Customers.** The individual-level parameters  $\{k, \lambda, \mu\}$  follow gamma distributions across customers independently

$$k \sim \text{Gamma}(t, \gamma); \quad (4)$$

$$\lambda \sim \text{Gamma}(r, \alpha); \quad (5)$$

$$\mu \sim \text{Gamma}(s, \beta). \quad (6)$$

Considering these adapted assumptions for the transaction process, we call this new model variant as the Pareto/GGG (gamma-gamma-gamma) model of repeat purchase behavior.

#### 3.2. Parameter Estimation

To achieve the parameter estimation for the Pareto/GGG, we formulate a full hierarchical Bayesian model with hyperpriors for the heterogeneity parameters, then generate draws of the marginal posterior distributions using a Markov Chain Monte Carlo (MCMC) sampling scheme. This comes with additional computational costs and implementation complexity, compared with the maximum likelihood method available for

Pareto/NBD, but we simultaneously gain the benefits of (1) estimated marginal posterior distributions rather than point estimates, (2) individual-level parameter estimates, and thus (3) straightforward simulations of customer-level metrics that are of managerial interest.

Rossi and Allenby (2003) provided a blueprint for applying a full Bayes approach (in contrast to an empirical Bayes approach) to hierarchical models such as Pareto/NBD. Then Ma and Liu (2007) published a specific MCMC scheme, comprised of Gibbs sampling with slice sampling to draw from the conditional distributions. Later Abe (2009) provided a significantly faster sampling scheme for the Pareto/NBD model by using data augmentation (Tanner and Wong 1987), thus expanding the parameter space with two latent variables: unobserved lifetime  $\tau$  and activity status  $z$ . This approach effectively decouples the sampling of the transaction process from the dropout process and therefore is able to take advantage of simple Bayesian updating rules for the conjugate priors.<sup>3</sup>

However, for Pareto/GGG, the transaction process priors are no longer conjugate priors, so we must sample the conditional posteriors for the individual-level transaction parameters  $k$  and  $\lambda$  using MCMC sampling for each Gibbs iteration and for each customer. The continued increases in computational power make even such brute-force simulation methods feasible and enable greater flexibility in model assumptions, as was also demonstrated by Abe's (2009) Pareto/NBD variant.

As we show subsequently, the Pareto/GGG requires only one additional, easily maintained summary statistic of historic transaction timings: the *sum over the logarithmic ITTs*.<sup>4</sup> Therefore, the data requirements imposed by modeling gamma-distributed ITTs are not, in practice, any higher than those for Pareto/NBD, which requires (1) the number of past transactions  $x$ , (2) the timing of the most recent transaction  $t_x$ , and (3) the total observation time  $T$  since the customer was acquired. Computing these three statistics requires processing the customer's full transaction history or their continuous updating whenever a new transaction is recorded. In either case, adding the sum of the logarithmic ITTs as an additional measure is straightforward to implement.

### 3.3. Key Expressions

Following, we present expressions for quantities of interest to users of the Pareto/GGG model. In doing

so, we use  $f_T$  to denote the density and  $F_T$  to indicate the cumulative distribution function of the gamma distribution.

*P(alive)*. The probability that a customer is still alive at time  $T$  is derived in Appendix A and results in the following equation:

$$P(\tau > T | k, \lambda, \mu, t_1, \dots, t_x, T) = \left( 1 + \frac{\int_{t_x}^T (1 - F_T(y - t_x | k, k\lambda)) \mu e^{-\mu y} dy}{(1 - F_T(T - t_x | k, k\lambda)) e^{-\mu T}} \right)^{-1}. \quad (7)$$

*Individual-level likelihood*. The likelihood of observing  $x$  intertransaction times  $\Delta t_j$  and then having no further transaction occur until time  $T$  (or in case of churn, until time  $\tau$ , i.e.,  $\Delta t_{x+1} > \min(T, \tau) - t_x$ ) can be expressed as follows:

$$\begin{aligned} L(k, \lambda | t_1, \dots, t_x, T, \tau) &= \left( \prod_{j=1}^x f_T(\Delta t_j | k, k\lambda) \right) (1 - F_T(\min(T, \tau) - t_x | k, k\lambda)) \\ &= \frac{(k\lambda)^{kx}}{\Gamma(k)^x} e^{-k\lambda t_x} \left( \prod_{j=1}^x (\Delta t_j)^{k-1} \right) \\ &\quad \cdot (1 - F_T(\min(T, \tau) - t_x | k, k\lambda)). \end{aligned} \quad (8)$$

*Conditional log-posterior for  $k$* . It follows from Equations (8) and (4) that

$$\begin{aligned} \log \pi(k | t_1, \dots, t_x, T, \tau, \lambda, t, \gamma) &\propto \log \text{likelihood} + \log \text{prior} \\ &\propto kx \log(k\lambda) - x \log \Gamma(k) - k\lambda t_x + (k-1) \sum_{j=1}^x \log \Delta t_j \\ &\quad + \log(1 - F_T(\min(T, \tau) - t_x | k, k\lambda)) \\ &\quad + (t-1) \log k - \gamma k. \end{aligned} \quad (9)$$

Note that the conditional log-posterior for the regularity parameter  $k$  requires the sum over the logarithmic ITTs as an additional summary statistic.

*Conditional log-posterior for  $\lambda$* . It follows from Equations (8) and (5) that

$$\begin{aligned} \log \pi(\lambda | t_1, \dots, t_x, T, \tau, k, r, \alpha) &\propto \log \text{likelihood} + \log \text{prior} \\ &\propto kx \log(\lambda) - k\lambda t_x + \log(1 - F_T(\min(T, \tau) - t_x | k, k\lambda)) \\ &\quad + (r-1) \log \lambda - \alpha \lambda. \end{aligned} \quad (10)$$

*Probability distribution of  $\tau$  in case of churn*. For sampling the lifetime of a customer who churns before  $T$ , we must consider the likelihood that no further transactions occur in  $(t_x, \tau]$ , such that the next ITT will be greater than  $\tau - t_x$ . The probability distribution of  $\tau - t_x$  thus can be specified up to a normalizing constant as follows:

$$f(\tau - t_x | k, \lambda, \mu) \propto e^{-\mu(\tau - t_x)} (1 - F_T(\tau - t_x | k, k\lambda)). \quad (11)$$

<sup>3</sup> Other studies using hierarchical Bayesian approaches using MCMC sampling to estimate the Pareto/NBD model or variations thereof include the contributions by Singh et al. (2009), Conoor (2010), or, more recently, Quintana and Marshall (2015) in a noncontractual setting, and the paper by Borle et al. (2008) in a contractual context.

<sup>4</sup> Note that one of the four clumpiness measures introduced by Zhang et al. (2013) relies on the same summary statistic, which again shows that the same underlying concept is being captured, just with opposite signs.

### 3.4. MCMC Procedure

The sampling scheme to generate draws from the joint posterior distribution is a combination of Ma and Liu's (2007) Gibbs sampler with slice sampling (Neal 2003) and Abe's (2009) augmented parameter space.

1. Use separate gamma distributions as hyperpriors for  $t$ ,  $\gamma$ ,  $r$ ,  $\alpha$ ,  $s$ , and  $\beta$ , and set the according shape and rate hyperparameters  $(t_1, t_2)$ ,  $(\gamma_1, \gamma_2)$ ,  $(r_1, r_2)$ ,  $(\alpha_1, \alpha_2)$ ,  $(s_1, s_2)$ , and  $(\beta_1, \beta_2)$  to capture the prior belief on the heterogeneity parameters.

2. Set initial values for  $t$ ,  $\gamma$ ,  $r$ ,  $\alpha$ ,  $s$ , and  $\beta$ , such as by using maximum likelihood estimates of Pareto/NBD.

3. Set initial values for  $\{k_i, \lambda_i, \mu_i, z_i, \tau_i\}$  for all customers.

4. For each customer  $i$ ,

(a) Draw  $k_i$  by slice sampling the conditional posterior in Equation (9);

(b) Draw  $\lambda_i$  by slice sampling the conditional posterior in Equation (10);

(c) Draw  $\mu_i \sim \text{Gamma}(s+1, \beta+\tau_i)$ , with the recognition that the gamma distribution is the conjugate prior of the exponential distribution;

(d) Draw  $z_i \sim \text{Bernoulli}(P(\text{alive}))$  with  $P(\text{alive})$  calculated according to Equation (7);

(e) Draw  $\tau_i$  conditional on status  $z_i$ :

(i) If the customer is alive ( $z_i = 1$ ), then draw  $\tau_i \sim \text{Exponential}(\mu_i)$ , left truncated to  $[T_i, \infty)$ ;

(ii) If the customer has already churned before time  $T_i$  ( $z_i = 0$ ), then draw  $\tau_i - t_{x_i}$  by slice-sampling the probability density in Equation (11), truncated to  $[0, T_i - t_{x_i}]$ .

5. Draw the heterogeneity parameters, treating the individual-level parameter draws as data and the specified gamma hyperpriors as priors. Ma and Liu (2007) propose updating the rate and shape heterogeneity parameters separately, whereas we suggest using component-wise slice-sampling to draw them simultaneously, which reduces the strong auto-correlation of the MCMC chain.

(a) Draw  $\pi(t, \gamma | \{k_i\})$ .

(b) Draw  $\pi(r, \alpha | \{\lambda_i\})$ .

(c) Draw  $\pi(s, \beta | \{\mu_i\})$ .

6. Repeat Steps 4 and 5 until convergence is reached and sufficient samples have been drawn.

## 4. Performance Evaluation and Empirical Analysis

We benchmark the performance of the Pareto/GGG against the Pareto/NBD using the following evaluation strategy: First, we conduct an extensive simulation study to systematically investigate the role of ITT regularity in holdout-forecasting tasks across variations of our model's assumptions. Second, we assess the empirical performance of the model using six real-world data sets on the purchasing of various product categories at

e-commerce websites, an online grocery retailer, and the donation records of a nonprofit organization.

### 4.1. Simulation Study

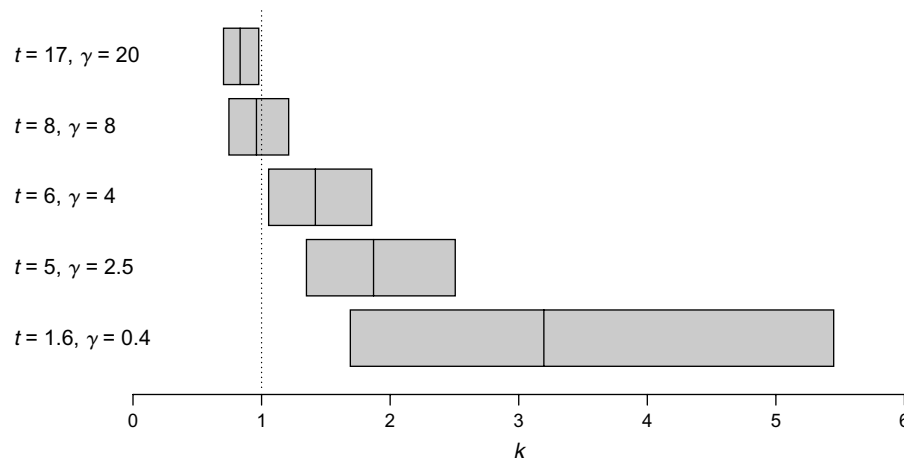
Our simulation study sought a better understanding of the benefits of incorporating regularity in a wide variety of purchase settings. For this purpose, we built on the simulation design suggested by Fader et al. (2005a), who use three levels for each of the four Pareto/NBD heterogeneity parameters for creating synthetic cohorts. To ensure a reasonable size for the total number of simulated "worlds," we chose only the two extreme values for each parameter, but combined these  $2^4 = 16$  settings with five distributions of regularity and two cohort sizes, resulting in a total full-factorial design of  $16 \times 5 \times 2 = 160$  Pareto/GGG scenarios. The chosen parameter values are as follows:  $N \in \{1,000, 4,000\}$ ,  $r \in \{0.25, 0.75\}$ ,  $\alpha \in \{5, 15\}$ ,  $s \in \{0.25, 0.75\}$ ,  $\beta \in \{5, 15\}$ , and  $(t, \gamma) \in \{(1.6, 0.4), (5, 2.5), (6, 4), (8, 8), (17, 20)\}$ . The resulting distributions for the regularity parameter  $k$  are displayed in Figure 3; they include a mix of customers with clumpy, random, and regular transaction timing. Based on these assumptions, we then generated transaction records for a calibration period and a holdout period of 52 weeks each. Similar to the simulation environment created by Fader et al. (2005a), the spanned parameter space covers a wide range of settings. The share of customers with no repeat transactions during the calibration period ( $x = 0$ ) ranges from 22% to 91%; the share of frequent customers with  $x \geq 10$  spans 0% to 22%; and the share of customers who are active during the holdout period ranges from 4% to 58%.

For each scenario, we performed parameter estimations using the MCMC sampler, with weakly informative hyperpriors. To ensure like-for-like comparisons between the Pareto/GGG and Pareto/NBD, we also performed parameter estimations for the latter using MCMC sampling. An efficient implementation of the Pareto/GGG and Pareto/NBD MCMC sampler is made available as part of the BTYDplus R package (R Core Team 2014, Eddelbuettel and François 2011) under an open-source license.<sup>5</sup> Further details on runtime and convergence diagnostics are reported in Appendix B.

To illustrate parameter recovery, Table 1 shows the results for five selected scenarios. Apparently, the Pareto/GGG MCMC sampler can recover the underlying data-generating parameters quite well, and it does so more effectively for the purchase process  $(t, \gamma, r, \alpha)$  than for the unobservable lifetime process  $(s, \beta)$ . A further analysis of all 160 scenarios confirms

<sup>5</sup> The authors wish to express their gratitude to Sandeep Conoor, who provided them with a working Fortran implementation of his MCMC sampler for estimating the Pareto/NBD. The work by Conoor (2010) proved to be very helpful for writing our own performance-tuned MCMC sampler in R.



**Figure 3** Interquartile Ranges for Simulated Distributions of Regularity Parameter  $k$ 

that Pareto/NBD tends to overestimate lifetime and ITTs in the presence of regularity. This error likely results because the Pareto/NBD model interprets a long transaction hiatus observed for a regular customer rather as an exceptionally long ITT, because it allows for greater variance in waiting times than would be the case when taking regularity into account.

Next, we assess the impact of incorporating regularity into the model by examining the lift in predictive accuracy for all 160 simulated scenarios when we move from Pareto/NBD to Pareto/GGG. Forecasting accuracy is compared in terms of the customer-level mean absolute error (MAE) of the predicted number of transactions during the holdout period;<sup>6</sup> we define a relative lift as  $1 - \text{MAE}_{\text{PGGG}} / \text{MAE}_{\text{PNBD}}$  and an absolute lift as  $\text{MAE}_{\text{PNBD}} - \text{MAE}_{\text{PGGG}}$ . Thus, the higher the lift measure, the higher the (relative) gain in predictive accuracy. In general, the Pareto/GGG performs well in these forecasting tasks. In cases with predominantly random or clumpy ITT patterns, the models generally perform on par. However, for scenarios with mildly regular timing patterns, the Pareto/GGG already consistently improves forecasting accuracy across the board, with larger improvements for greater degrees of regularity (up to +20% relative lift). As expected, the stronger the regularity within a cohort, the stronger the lift, and this result holds across all other parameter configurations. A complete summary of the results of the simulation study for all 160 synthetic scenarios is included in Appendix B.

To obtain a more thorough understanding of the particular customer groups within a cohort for which the largest gains in predictive accuracy can be achieved

when accounting for regularity, we combined the forecasts for all 400,000 customers from the 160 simulated worlds and then divided them into distinct segments according to their recency, frequency, and (true underlying) regularity. In terms of frequency, we distinguish groups of customers with four or more transactions (high), one to three transactions (low), and no repeat transactions (zero). For recency, the distinction indicates whether a customer conducted the latest transaction less than eight weeks ago,  $t_x > 42$  (high), or more than eight weeks ago (low). Table 2 reports the relative and absolute lift in MAE both for Pareto/NBD and Pareto/GGG, along with the average number of transactions during the holdout period for that segment. We also rank our 400,000 synthetic customers according to their regularity, divide them into 10 equally sized groups, and plot their corresponding relative lift in MAE against their mean regularity in Figure 4. Several important findings emerge from inspecting Table 2 and Figure 4:

- The stronger the regularity, the greater the lift in the predictive accuracy of the Pareto/GGG compared with the Pareto/NBD forecasts.
- For customers with random ( $k \approx 1$ ) or even clumpy ( $k < 1$ ) purchase patterns, the Pareto/GGG offers predictions that are generally on par with those of the Pareto/NBD.
- The lift for customers purchasing at high frequencies is greater than for customers with lower purchase frequencies, likely because Pareto/GGG detects an individual's degree of regularity more easily when more transactions are observed in the past. However, we find a lift even for customers with few or zero repurchase transactions. In such cases, the model leverages the estimated heterogeneity of regularity to form a prior belief about each person's regularity.
- The lift for customers with a longer purchase hiatus since the last observed purchase (i.e., the low recency group) is greater than that for those who were

<sup>6</sup> The MAE measure is frequently used for time-series data. In the context of customer-base analysis, see the recent contribution by Schwartz et al. (2014) for a justification of choosing MAE as a model selection criterion.

**Table 1 Recoverability of True Parameters for Five Selected Scenarios with  $N = 4,000$**

	$t_{q50}$	$\gamma_{q50}$	$r_{q50}$	$\alpha_{q50}$	$s_{q50}$	$\beta_{q50}$	$k_{q10}$	$k_{q50}$	$k_{q90}$	$ITT_{q50}$	$\tau_{q50}$	$P(\text{alive})$ (%)
True	1.6	0.4	0.25	5.0	0.25	5.0	0.84	3.21	8.22	91	75	54
PGGG	1.9	0.5	0.22	4.4	0.31	10.9	0.94	3.11	7.48	129	83	57
PNBD			0.18	4.3	0.27	14.8				205	183	67
True	5.0	2.5	0.75	5.0	0.75	5.0	0.97	1.87	3.20	9	8	16
PGGG	4.8	2.4	0.65	4.5	0.83	6.7	0.95	1.84	3.19	10	9	17
PNBD			0.54	4.5	0.79	7.6				12	11	20
True	6.0	4.0	0.25	5.0	0.25	15.0	0.79	1.42	2.32	82	224	69
PGGG	5.5	3.6	0.23	4.4	0.33	31.5	0.77	1.43	2.39	96	229	73
PNBD			0.21	4.3	0.35	53.4				111	296	78
True	8.0	8.0	0.25	5.0	0.75	15.0	0.58	0.96	1.47	71	23	33
PGGG	7.3	7.4	0.24	4.6	0.72	13.8	0.55	0.94	1.47	71	22	32
PNBD			0.30	4.7	0.54	5.3				44	14	28
True	17.0	20.0	0.75	15.0	0.75	15.0	0.60	0.83	1.12	21	23	33
PGGG	14.4	17.1	0.72	14.9	0.80	17.1	0.58	0.83	1.15	22	24	33
PNBD			0.94	15.3	0.65	8.4				17	15	27

Notes. PNBD, Pareto/NBD; PGGG, Pareto/GGG.

active recently. It seems (and we will subsequently confirm this) that the presence of regularity particularly facilitates distinguishing between active and inactive customers, if their next transaction is overdue. While this finding makes intuitive sense, traditional NBD-type models fail to take advantage of it. If the timing patterns are fairly erratic, this becomes relatively inconsequential. However, in a world with increasingly regular purchase timing, the Pareto/NBD no longer provides good predictions and is clearly outperformed by the much more flexible Pareto/GGG.

- The greatest lift emerges for the group of customers who formerly purchased very frequently in the past but have not been active more recently. Although

usually a relatively small segment, it deserves particular attention by managers because these valuable customers are currently at risk of being lost. Accounting for regularity helps to remove some of the ambiguity regarding their future behavior.

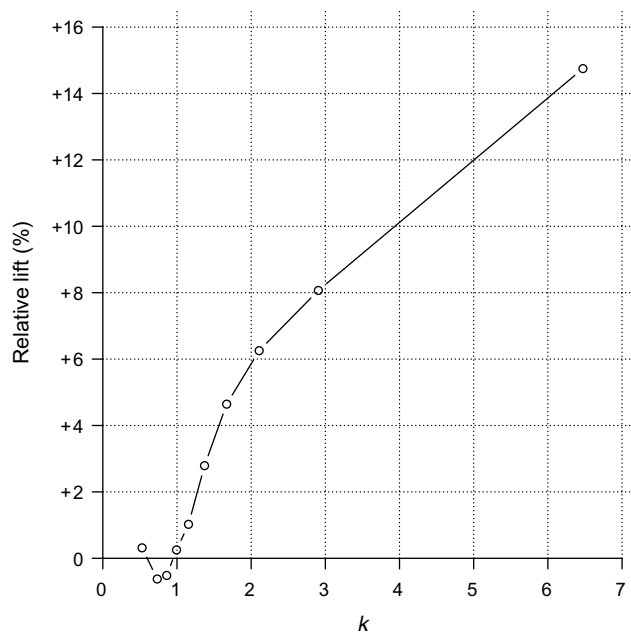
- Finally, note that for the high-frequency, low-recency segment, the mean number of transactions during the holdout period  $x^*$  is significantly lower for regular than for random customers (2.31 vs. 3.49; see Table 2). A purely RF-based model would not be able to capture such a pattern.

These findings suggest refining the observation of Zhang et al. (2015, p. 206) that “a buy-till-you-die story performs well for nonclumpy customers, but

**Table 2 Impact of Incorporating Regularity by Customer Segment**

Customer segment			Lift		MAE		Holdout mean( $x^*$ )
Regularity	Recency	Frequency	Relative (%)	Absolute	PGGG	PNBD	
None $k < 1.5$	Zero	Zero	<b>+0</b>	+0.00	0.16	0.16	0.09
	Low	Low	<b>+2</b>	+0.02	0.87	0.89	0.81
	Low	High	<b>−0</b>	−0.00	2.64	2.64	3.49
	High	Low	<b>+1</b>	+0.01	1.69	1.71	2.24
	High	High	<b>+0</b>	+0.00	4.83	4.83	11.49
	All	All	<b>+0</b>	+0.00	0.87	0.88	1.35
Low $1.5 < k < 3$	Zero	Zero	<b>+11</b>	+0.02	0.14	0.16	0.08
	Low	Low	<b>+9</b>	+0.08	0.83	0.91	0.98
	Low	High	<b>+15</b>	+0.34	1.97	2.31	3.01
	High	Low	<b>+1</b>	+0.02	1.28	1.30	2.21
	High	High	<b>+1</b>	+0.06	3.81	3.87	11.44
	All	All	<b>+6</b>	+0.05	0.69	0.74	1.36
High $k > 3$	Zero	Zero	<b>+16</b>	+0.02	0.13	0.15	0.07
	Low	Low	<b>+18</b>	+0.16	0.73	0.89	1.05
	Low	High	<b>+41</b>	+0.94	1.33	2.27	2.31
	High	Low	<b>+4</b>	+0.04	0.97	1.01	2.08
	High	High	<b>+7</b>	+0.25	3.09	3.33	11.61
	All	All	<b>+15</b>	+0.10	0.55	0.65	1.33

Notes. PNBD, Pareto/NBD; PGGG, Pareto/GGG. Values in bold describe the relative lift in MAE for the Pareto/GGG against the Pareto/NBD.

**Figure 4** Relative Lift in MAE vs. Regularity  $k$ 

not for clumpy ones.” Although the clumpiness phenomenon might indeed be better accommodated in a modeling framework that allows for a more complex nonstationary repeat-buying behavior, we advocate replacing the dichotomy of clumpy versus nonclumpy with the understanding that timing patterns range along a continuum between clumpy and regular (with improved predictions for the latter when adopting the Pareto/GGG).

In noncontractual settings, the expected number of future transactions depends largely on the assessment of a customer’s latent status. The BTYD model class assumes that a customer who has defected is “lost for good” and will not make any further transactions in the future. To better understand the previous discussion on the varying impact of regularity for predicting future transactions (Table 2), a closer examination of the functional shape of  $P(\text{alive})$  (see §3.3) and its interplay with recency, frequency, and regularity is helpful. Figure 5 displays the dependence of  $P(\text{alive})$  on observed recency for two levels of frequency ( $\lambda \in \{1/6, 1/26\}$ ), for various degrees of ITT regularity ( $k \in \{0.5, 1, 2, 4, 8, \infty\}$ ), and for a hypothetical customer with a mean lifetime of 52 weeks ( $\mu = 1/52$ ). The thick solid curve ( $k = 1$ ) displays the corresponding functional shape of a Poisson purchase process, paired with an exponentially distributed lifetime (i.e., Pareto/NBD). The thin solid line ( $k = \infty$ ) instead shows the extreme case of equally spaced transaction timings. The dotted line represents a clumpy customer, and the three dashed lines represent various degrees of regularity. A closer inspection of Figure 5 reveals that the deviation from Pareto/NBD depends largely on whether the latest

transaction hiatus  $T - t_x$  is greater or smaller than the expected ITT. If the transaction is overdue (i.e.,  $T - t_x > \text{mean}(\text{ITT})$ ),  $P(\text{alive})$  declines for regular customers, and the magnitude of this shift depends on the strength of the regularity.

Our motivating example in Figure 1 reflects this finding: It indicates the timing patterns of customers A and B with the same recency and frequency but different degrees of regularity. Figure 5(a) also indicates their approximate positions on the corresponding  $P(\text{alive})$  curves. Their observed purchase hiatus of 14 weeks yields a 44% probability of being alive for customer A with random purchase occurrences; it is close to 0% for regular customer B. For the border case with deterministic timing patterns,  $P(\text{alive})$  falls to zero immediately after the (constant) ITT has elapsed without activity. In the case of clumpy patterns, with strongly varying ITTs, the effect moves in the other direction, resulting in greater uncertainty regarding the latent activity state of the customer. This is exactly reflected by the findings of Zhang et al. (2015) of larger prediction errors for clumpy customers using a BTYD model (in their case, the BG/BB by Fader et al. 2010). For customers who were recently active (again, in relation to their expected ITT), as depicted to the right of the curves’ inflection points, we find only marginal differences. Comparing Figures 5(a) and 5(b) further supports our previous finding from the simulation study that the largest gain in predictive accuracy should be expected for customers with high frequency and low recency because, for these segments, regularity allows us to remove some of the ambiguity concerning the customer’s status.

In sum, our simulation study shows that the presence of regularity allows for better predictability. Thus, replacing the NBD with the more flexible GGG-type transaction model is particularly advisable for data sets exhibiting regular timing patterns.

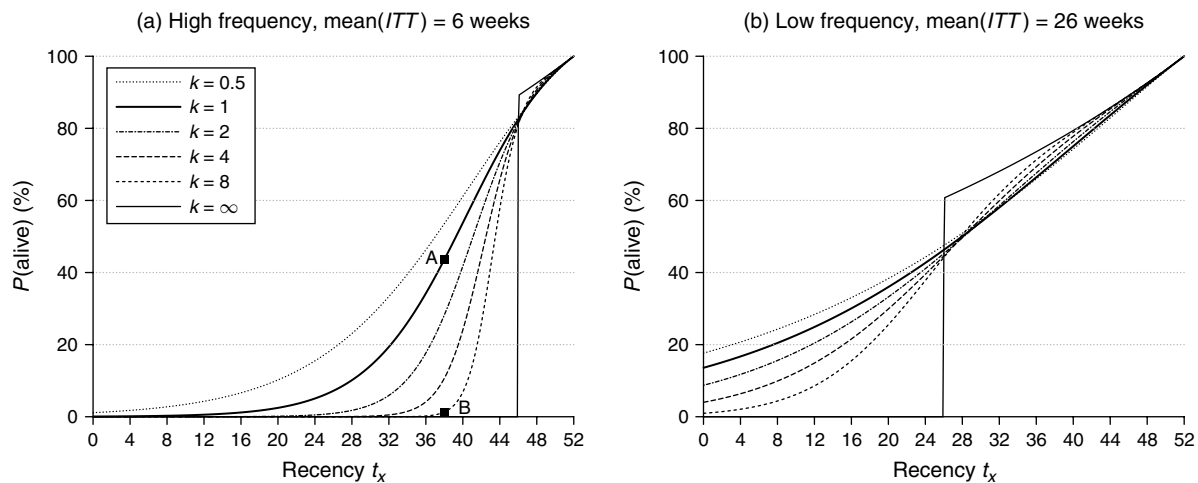
#### 4.2. Empirical Application of the Pareto/GGG Model

We now empirically examine the importance of incorporating ITT regularity into stochastic models, using six data sets that represent various settings.

- **CDNOW:** This data set includes 2,357 customers of an online CD store (CDNOW) who were acquired in the first quarter of 1997 and then observed over 1.5 years. This canonical data set has been studied and benchmarked extensively in the marketing literature (Fader and Hardie 2001; Fader et al. 2005a, b; Batislam et al. 2007; Wübben and von Wangenheim 2008; Abe 2009; Jerath et al. 2011; Bemmaror and Gladys 2012; Zhang et al. 2015).

- **Apparel and accessories:** This data set includes 831 customers of an online apparel and accessories retailer ([www.m18.com](http://www.m18.com)) who were acquired in April

Figure 5 Interplay of Recency, Frequency, Regularity, and  $P(\text{alive})$



2009 and observed over the course of one year. This data set came from Zhang et al. (2015), who kindly provided us with access to their data.

- **Donations:** This data set includes 21,166 donors to a nonprofit organization who were acquired in the first half of 2002 and observed over 4.5 years. The data set was provided by the Direct Marketing Educational Foundation (see also Malthouse 2009, Bemmar and Gladly 2012, Schweidel and Knox 2013).

- **Groceries:** These data came from an online retailer offering a broad range of grocery categories. The total observation period spans four years. Because customers' acquisition date was not part of this data set, we constructed a quasi cohort by limiting the analysis to the 1,525 customers who purchased in the first quarter of 2006 but not at all in the preceding two years.<sup>7</sup> These purchase data are available at the product category level, which allows us to study purchase patterns by category.

- **Dietary supplements and office supplies:** Two additional data sets come from an anonymous e-commerce service provider, each consisting of 35 weeks of purchase records for 1,000 randomly sampled customers.

All data sets contain the complete transaction records, including exact dates. Repeated transactions by a customer on a given day are treated as a single transaction. Without loss of generality, the chosen unit of time for our analysis is one week, and hence we provide the reported lifetime and ITT estimates in weeks.<sup>8</sup> Figure 6 provides an overview of all six data sets by displaying the timing patterns of 40 randomly sampled customers,

plus the histogram of transaction counts divided by calibration and holdout period. The displayed timing patterns reveal the sparseness of the available customer-level information, which requires us to pool data across customers. Furthermore, we can visually detect the varying degrees of regularity across not only the data sets but also the frequent customers in the databases. We seek to quantify this regularity and its heterogeneity by fitting the Pareto/GGG model.

For each data set, we fit both the Pareto/GGG and Pareto/NBD models using MCMC sampling, with the same settings as in the simulation study: four chains with 8,000 samples, of which the initial 2,000 are the burn-in sample. For the donations data set, we increased the samples to 30,000 because of the high autocorrelation in the draws for the lifetime parameters  $s$  and  $\beta$ .

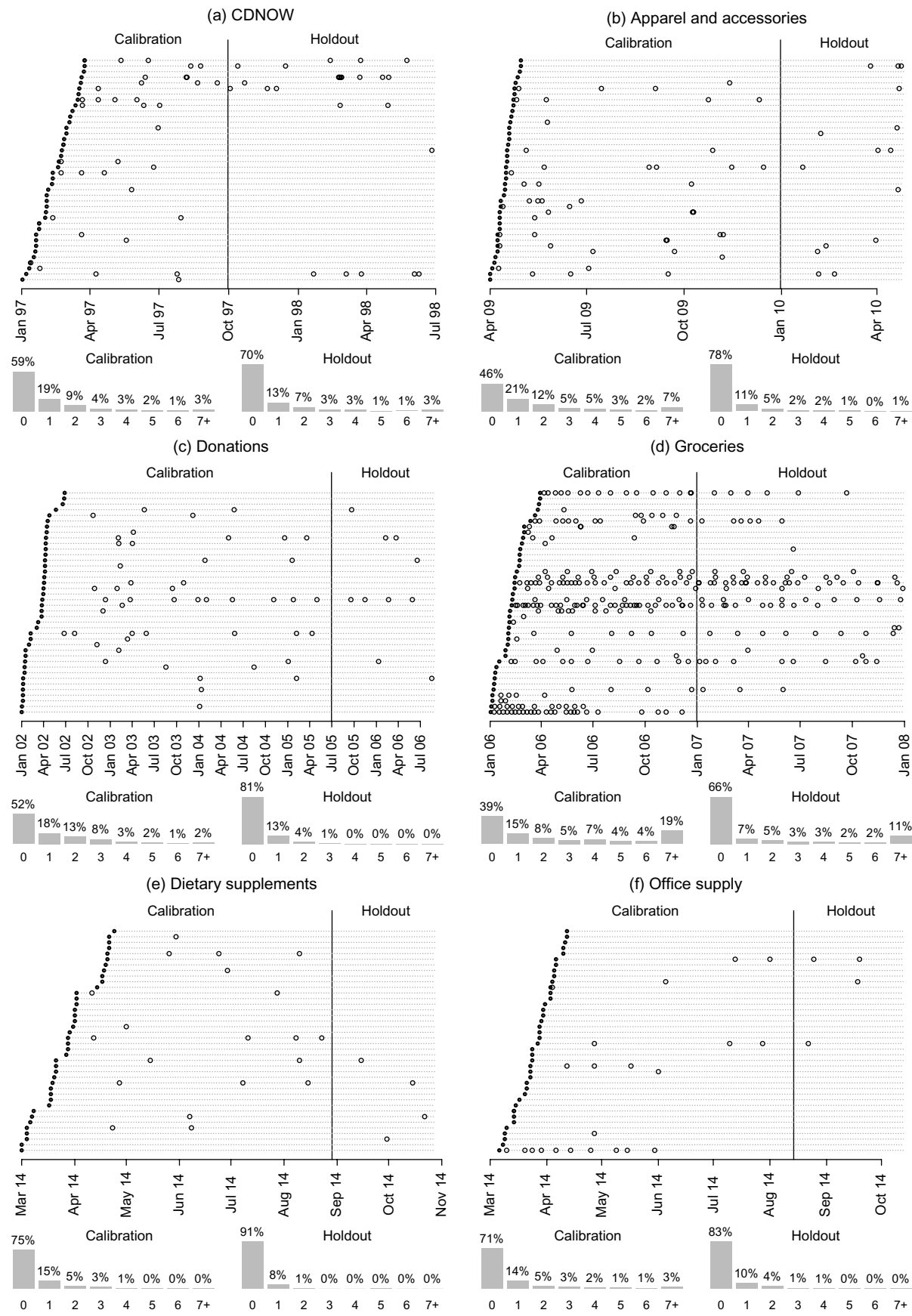
A summary overview of the resulting posterior distributions for the model parameters is given in Table 3; the ranges of the respective posterior estimates for regularity are depicted in Figure 7. For the CDNOW data set, the regularity parameter  $k$  varies narrowly around 1, which confirms the validity of assuming a Poisson purchase process for this specific customer base. The estimates for the remaining Pareto/GGG parameters closely match those for Pareto/NBD, and both models result in similar fit and predictions. For the apparel and accessories retailer, the estimates for  $k$  also exhibit little variation, although the timing patterns appear rather irregular ( $k_{q50} = 0.85$ ). For the remaining four data sets, we detected varying degrees of regularity, with a median  $k$  ranging from 1.78 for dietary supplements to 3.47 for the grocery retailer (see Table 3). However, as can also be seen from Figure 7, the individual-level estimates of the regularity parameter  $k$  vary significantly within these customer bases. In the grocery data set, for example, close to 10% of customers are estimated to purchase with timing patterns that are

<sup>7</sup> Batislam et al. (2007) use this approach by left filtering a grocery retail customer base with the requirement that it provides 11 months of initial inactivity.

<sup>8</sup> Note that the chosen unit of time is an arbitrary definition for continuous-time BTYD models. It is only reflected in the scale parameters of the purchase and dropout models, but does not impact the predictions themselves.



Figure 6    **Sampled Timing Patterns and Histograms of Transaction Counts**



**Table 3** Selected Quantiles of the Posterior Distributions for Pareto/GGG and Pareto/NBD Parameters

		$t_{q50}$	$\gamma_{q50}$	$r_{q50}$	$\alpha_{q50}$	$s_{q50}$	$\beta_{q50}$	$k_{q10}$	$k_{q50}$	$k_{q90}$	$ITT_{q50}$	$\tau_{q50}$	$P(\text{alive})$ (%)
CDNOW	PGGG	40.4	38.9	0.5	10.3	0.6	11.5	0.86	1.04	1.29	26.2	24.1	44
	PNBD			0.6	10.6	0.5	8.3				25.8	25.2	45
Apparel and accessories	PGGG	43.9	51.0	0.5	7.5	0.6	19.1	0.69	0.85	1.03	18.0	42.1	53
	PNBD			0.6	7.4	0.6	15.0				16.1	32.6	47
Donations	PGGG	4.2	1.4	1.0	77.0	8.2	1,565	1.33	2.70	4.88	96.4	136.4	42
	PNBD			0.6	77.4	2.0	1,432				158.3	628.7	81
Groceries	PGGG	1.7	0.4	0.9	5.1	0.4	4.5	1.02	3.47	9.24	7.1	17.7	35
	PNBD			0.8	5.7	0.4	5.5				8.0	28.7	43
Dietary supplements	PGGG	5.8	3.0	0.7	17.7	0.5	5.5	1.00	1.78	3.20	38.9	19.1	48
	PNBD			0.5	20.5	0.3	12.0				70.5	99.1	71
Office supply	PGGG	4.8	2.1	0.3	4.3	0.4	6.2	1.07	2.08	3.66	48.8	34.6	59
	PNBD			0.2	4.5	0.3	20.7				95.6	159.5	80

Notes. PNBD, Pareto/NBD; PGGG, Pareto/GGG. Values in bold describe the relative lift in MAE for the Pareto/GGG against the Pareto/NBD.

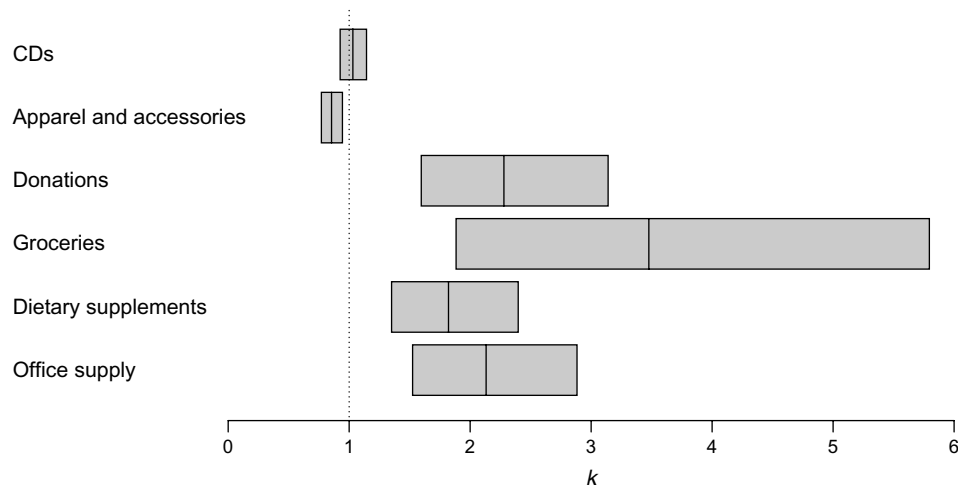
more irregular than random ( $k < 1$ ), and another 10% exhibit strong regular patterns with  $k$  larger than 9.24 (see the corresponding quantiles for  $k_{q10}$  and  $k_{q90}$  in Table 3).

We also investigated how many of the individual-level marginal posterior densities for which we have 90% confidence that the timing patterns are more regular than random (i.e.,  $P(k > 1) > 0.9$ ). Whereas for the CDNOW and the apparel and accessories data sets there were no such customers, the share of customers who satisfy this condition is 95% for donations, 72% for groceries, 58% for dietary supplements, and 80% for the office supply data set. This suggests that regularity is a widely prevalent phenomenon at least for some of the empirical data we explored and for substantial fractions of customers within these data sets. Our observation is perfectly consistent with those from the related literature reviewed by §2 and with the findings by Zhang et al. (2015), who report the presence of clumpy timing patterns to be particularly present in online visitations and to a far lesser extent for repeated

purchases. It is an interesting research subject to study whether these findings also hold for a broader set of empirical settings. Furthermore, Table 3 shows that for data sets with mainly regular patterns, the Pareto/NBD results in significantly higher estimates for lifetime  $\tau$ , and thus  $P(\text{alive})$ , compared to Pareto/GGG. This finding accords with those from the simulation study, which diagnosed a systematic bias for the Pareto/NBD in the presence of regularity.

Corresponding to the simulation study, we also assess forecasting accuracy using the MAE of the predicted number of transactions during the holdout period and the comparative lift when we incorporate regularity. In addition, we provide the MAE for a simple heuristic forecast following a method suggested by Wübben and von Wangenheim (2008). Table 4 reports the results for all six empirical data sets, which conform to those of the simulation study: the stronger the ITT regularity, the greater the lift in predictive accuracy, whereas for cases with predominantly random purchase occurrence, the models perform equally well. Note

**Figure 7** Interquartile Ranges for Posterior Distributions of Regularity Parameter  $k$



**Table 4** Impact of Incorporating Regularity by Data Set

	$\hat{k}_{\text{wheat}}$	$k_{q50}$	Lift		Mean absolute error			Holdout mean( $x^*$ )
			Relative (%)	Absolute	PGGG	PNBD	Heuristic	
CDNOW	1.0	1.0	<b>+2</b>	+0.02	0.76	0.77	1.02	0.80
Apparel and accessories	0.6	0.8	<b>−2</b>	−0.01	0.44	0.43	0.56	0.38
Donations	2.2	2.7	<b>+16</b>	+0.06	0.29	0.35	0.34	0.28
Groceries	2.5	3.4	<b>+8</b>	+0.13	1.39	1.52	2.57	2.22
Dietary supplements	2.0	1.8	<b>+5</b>	+0.01	0.15	0.16	0.15	0.09
Office supply	1.8	2.1	<b>+4</b>	+0.01	0.28	0.30	0.30	0.32

Notes. PNBD, Pareto/NBD; PGGG, Pareto/GGG. Values in bold describe the relative lift in MAE for the Pareto/GGG against the Pareto/NBD.

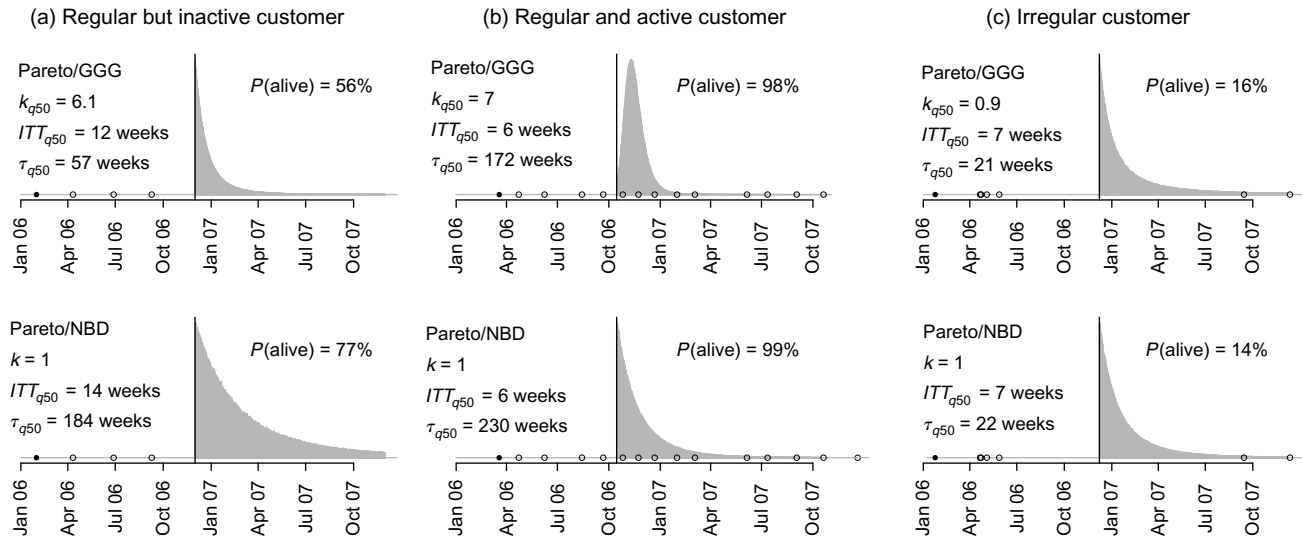
that by taking regularity into account, the probabilistic model outperforms the heuristic in all six cases. Next to the median of the posterior for the Pareto/GGG regularity parameter  $k$ , Table 4 also reports Wheat and Morrison's (1990) summary statistic calculated according to Equation (1). Except for the grocery data set, which is characterized by a considerable degree of heterogeneity,  $\hat{k}_{\text{wheat}}$  approximates the regularity estimated by the Pareto/GGG fairly well. Given our empirical findings and the diagnostic power of the Wheat and Morrison (1990) summary statistic, we expect that customer-base analysts would benefit from the improved performance of the Pareto/GGG relative to the Pareto/NBD when  $\hat{k}_{\text{wheat}}$  is estimated to be approximately 1.5 or higher. Note, however, that this is a data-set-level statistic and ignores any potential heterogeneity across individual customers.

To illustrate the changes in the individual-level estimates when we account for regularity, we inspect three selected customers from the grocery data set in detail. Figure 8 displays the timing patterns during the calibration and holdout periods, together with their corresponding Pareto/GGG and Pareto/NBD parameter estimates. Customer (a) engaged in four rather regularly spaced transactions during the calibration period, but was not active in the last months of 2006 (the end of the calibration period). The Pareto/GGG estimates a strong degree of regularity ( $k_{q50} = 6.1$ ) for this customer and assigns a probability of only 56% that this customer will purchase again, compared with the significantly higher probability of 77% according to the Pareto/NBD model. Note that the posterior probability density for the next transaction arrival shows a steep decline, implying that the next purchase event is overdue. To marketing managers, these are clear signals that the customer (rightly) is at risk of being lost and that appropriate marketing actions are called for. The Pareto/GGG performs substantially better at identifying this than does the Pareto/NBD, for which the long transaction hiatus is not as strong an indicator of defection. Customer (b) exhibits similarly strong regularity, but has remained active recently. Because of the recent transaction,  $P(\text{alive})$  is (correctly)

assessed equally high by the two models. However, the posterior probability density estimated by the Pareto/GGG points to an expected short inactivity period, after which transactions are expected to fall in a rather narrow bandwidth. Customer (c) undertook five rather irregularly spaced transactions in 2006, resulting in an estimate of  $k_{q50} = 0.9$ , which further demonstrates the variety of timing patterns that can appear in one and the same customer cohort. Because of the clumpy ITT pattern,  $P(\text{alive})$  is also slightly elevated when it is taken into consideration. In sum, these findings reinforce the previously discussed model behavior, as depicted in Figure 5. Apparently, the clumpy case is difficult to predict for both models, and the Pareto/NBD assumptions appear to be not particularly costly relative to the Pareto/GGG.

To conclude our empirical application, we exploit the shopping basket data at the product category level in the grocery data set to provide some further insights into which categories exhibit stronger regularities than others. For each of the 143 product categories, we construct quasi cohorts of customers who did not purchase in that category in 2004 and 2005 but did so in the first half of 2006. Using these cohorts, we then fit a Pareto/GGG model using all of 2006 as a calibration period, obtaining estimated distributions for  $k$  in each category. Among the most regularly bought categories, we find perishable food categories such as salad ( $k_{q50} = 7.7$ ) and fresh cheese ( $k_{q50} = 6.0$ ), as well as packaged, regularly consumed goods, such as washing detergents ( $k_{q50} = 6.0$ ), fabric conditioners ( $k_{q50} = 5.0$ ), aluminum foil ( $k_{q50} = 4.7$ ), ground coffee ( $k_{q50} = 4.2$ ), and toilet paper ( $k_{q50} = 3.4$ ). The less regular categories still exhibit median  $k$  values greater than 1.5, as exemplified by categories such as pantyhose ( $k_{q50} = 1.5$ ), spices ( $k_{q50} = 1.9$ ), and flour ( $k_{q50} = 2.0$ ). A benchmark of the Pareto/GGG against the Pareto/NBD further showed that for 140 of the 143 available categories, we increased predictive accuracy by accounting for observed regularity (the detailed results are available on request). As our findings suggest, ITT regularity generally translates into lower prediction errors. Thus, even for customers exhibiting considerable uncertainty in their future purchase behavior at the firm-level,

**Figure 8** Selected Grocery Customers and Their Posterior Probability of Next Transaction Timing



multicategory firms such as the grocery retailer in the above example could greatly benefit from leveraging information on category-level regularities in their overall assessment of customer activities.

## 5. Relationship Between the Regularity Parameter $k$ and the Clumpiness Metric $C$

Zhang et al. (2013) introduce a class of measures to capture clumpiness in incidence data. A subsequent contribution by Zhang et al. (2015) demonstrates the application of these measures to customer-base analysis using a specific  $C$  metric (i.e., the  $H_p$  variant) for a variety of purchase (purchase  $C$ ) and online visit (visit  $C$ ) patterns. The main empirical findings related to our research are that (1) clumpiness is a prevalent phenomenon, but mainly in the context of online visits and digital media consumption; (2) the inclusion of the  $C$  metric improves the data-fitting capability of RFM-based regression models of customers' future (i.e., out-of-sample) activity; and (3) clumpy customers tend to be more active than regular ones in future periods.<sup>9</sup> Our presented research supports these findings, provides a link to the theoretical framework underlying the well-established class of BTYD models and thus contributes to improving our understanding of them, and develops a predictive model capable of extrapolating beyond the calibration period.

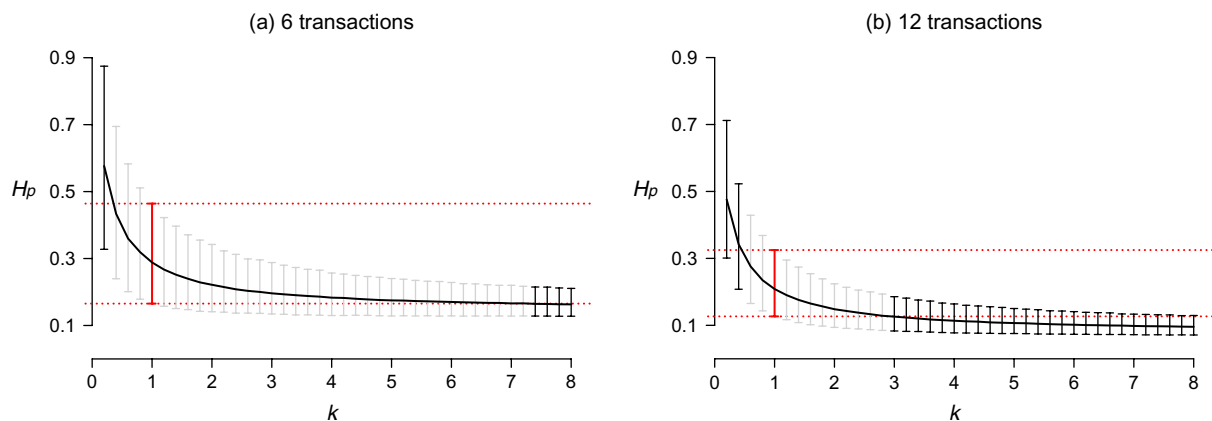
To establish the relationship between the Zhang et al. (2015)  $C$  metric and the shape parameter  $k$  of the gamma-timing model of the Pareto/GGG, we conducted another simulation study. For various values

of  $k$ , we generated 10,000 Gamma( $k, k\lambda$ ) timing patterns with  $n = 6$  and  $n = 12$  events each and calculated their  $H_p$  metric accordingly. The parameter  $\lambda$  is chosen sufficiently small to avoid zero-length ITTs when converting to discrete time units. The solid black curves in Figure 9 visualize the median over the  $H_p$  samples and reveal a strictly monotonous relationship between  $C$  and  $k$ : the higher  $k$  is, and thus the less variation we have in the ITTs, the lower the  $C$ -measure will be. Thus, despite being designed by the authors to measure clumpiness, the metric  $C$  also captures the degree of regularity.

Figure 9 further shows, for increments of  $k$ , the variation in the measure  $C$  as vertical lines, with the sampled 5% and 95% quantiles indicated by whiskers. Comparing Figure 9(a) with Figure 9(b) also demonstrates that the more transactions we observe per customer, the more confident we can be regarding the degree of clumpiness or regularity, respectively. In both graphs, there is significant overlap between the sampled  $C$  values across different timing patterns, but this is particularly pronounced for a smaller number of transactions. Furthermore, we extend the whiskers horizontally for the case of a simulated Poisson process (i.e.,  $k = 1$ ) by dotted lines, as these boundaries serve as the rejection regions for detecting clumpiness as described in Zhang et al. (2013). We find that only in cases of very strong clumpiness or strong regularity (marked by the whiskers in bold), the  $C$  measure is able to correctly reject the null hypothesis of randomly distributed events for more than half of the customers. Therefore, in settings characterized by customers with a small number of events (e.g.,  $n < 10$ ) during the observation period, analysts should be cautious when calculating the  $C$  measure, because it bears a significant

<sup>9</sup> Note the positive sign of the regression coefficient of purchase  $C$  in Table 4 reported by Zhang et al. (2015).



**Figure 9** (Color online) Calculations of Metric  $C$  for the Range of  $\text{Gamma}(k, k\lambda)$  Distributed Events

amount of uncertainty. The proposed Pareto/GGG addresses this problem in a Bayesian manner by first forming a prior belief regarding the timing patterns based on all customers and subsequently gradually updating this belief with each additional transaction at the customer level.

Another potential shortcoming of the  $C$  measure is its insensitivity in distinguishing between clumpy interevent timing and latent defection or churn. Any defected customer with a prolonged phase of inactivity at the end of the observation period will be biased toward clumpiness. This can also be seen from the evolution of the  $C$  measure for the third sample customer in Figure 2 of Zhang et al. (2013). When simulating an entire customer base following the Pareto/NBD assumptions, whereas parameters  $r$ ,  $\alpha$ ,  $s$ , and  $\beta$  are set to match the CDNOW parameter estimates (Fader et al. 2005a), 10% of the customer base will be incorrectly classified as clumpy, despite the rejection rate of the test being set to 5%. In these cases, the events might mistakenly appear to be “clustered” due to a long period of inactivity after a customer has dropped out. For churn settings with higher frequency ( $r = 0.75$ ,  $\alpha = 5$ ,  $s = 0.75$ ,  $\beta = 5$ ,  $T = 52$ ), the type I error can be as high as 35%.

However, we need to concede here that Zhang et al. (2015) primarily focus on visit  $C$ , and in digital settings such as those studied by the authors, infrequent visits and churn might not be a substantial issue. Thus, in such (or similar) settings, the  $C$  metric actually might be a useful tool for scanning a data set before any formal model fitting. However, in the case of purchase histories with significant shares of customers who purchase less frequently ( $n < 10$ ) and/or where customer defection is prevalent, the Pareto/GGG offers a descriptive and predictive alternative that is capable of avoiding both of the above-described shortcomings of the  $C$  measure.

## 6. General Discussion and Future Research

Many companies exploit the continuous influx of transaction data to make inferences about the future activity

of their customers and related metrics, such as CLV and its subcomponents (Fader et al. 2005a). Yet, despite the considerable scale of data available at the overall level, little information is typically available at the individual level because a significant share of customers engages in few (if any) repeat transactions. Thus, it is common practice to pool information across customers, and probabilistic purchase models based on RF data remain the primary means for doing so (Schwartz et al. 2014).

However, by condensing historic transaction records to RF summary statistics alone, these models discard any additional customer-level information that is contained in the past timing patterns but equally easily available to the analyst. Recently, Zhang et al. (2015) questioned whether recency and frequency are sufficient statistics to fully summarize a customer history and posit that adding an individual-level statistic reflecting a customer’s interevent timing patterns helps to better understand CLV and its subcomponents. The authors propose a metric-based approach to extend the widely adopted RF framework, with a measure to capture clumpiness in timing patterns, and they also demonstrate its usefulness for predicting customer value. Whereas Zhang et al. (2015) are very clear in positioning their work as a “measurement paper,” our contribution is a “modeling paper.” We complement their timely and important research by introducing a model-based approach, which accommodates a wide range of timing patterns (regular, random, and irregular) but adaptively pools the information across customers to attain reliable, individual-level estimates. A probabilistic modeling approach to capture regularity is not new to marketing (Chatfield and Goodhardt 1973, Morrison and Schmittlein 1988), but this research and the proposed Pareto/GGG model make two novel contributions. First, we propose adaptively pooling the sparse, individual-level information on timing patterns across customers and then leveraging heterogeneity in regularity to predict future behavior. Second, we build intuition and consistently demonstrate using an extensive simulation study and empirical applications

that, particularly in a BTYD setting, accounting for regularity is highly beneficial because it facilitates making inferences regarding the customer's latent status. In such settings, we gain the most in terms of predictive accuracy for the managerially important segment of highly frequent but recently inactive customers. All of these benefits entail only marginal additional costs in terms of data requirements, which involve simple and sufficient summary statistics of historical transaction records.

Beyond developing the Pareto/GGG model and comparing it to the Pareto/NBD model, our research in turn offers several important managerial implications. First, better predictions of important CLV components are of value to customer relationship managers per se because such improved estimates help them to prioritize customers more effectively. Second, the model results in sound customer-level estimates of the regularity parameter  $k$  and thus provide a valuable customer metric (as illustrated by Figure 8). For example, marketing managers can use this diagnostic information as a basis for segmenting customers (e.g., into irregular, random, and regular segments), as well as to score them according to their attractiveness and predictability. The corresponding  $R$ - and  $F$ -related metrics could define further subsegments. Third, managers might apply this segmentation effectively in their customer targeting and resource allocation decisions. For example, regular customer types could be targeted with cross- or upselling options prior to their next projected visit; "overdue" regular customers should be gently reminded to return or solicited to provide customer feedback. Finally, companies operating in multiple categories could also learn from category-specific purchase timings to draw inferences concerning the (overall) activity status of their customers. For example, even customers showing random (or clumpy) ITT patterns at the firm level could reveal some aspects of regularity at the category level when examining their shopping baskets in greater detail, i.e., categories they purchase on a more regular basis. Indeed, our empirical study using online grocery data showed that there is considerable variation across categories. Model builders could leverage this information by extending our approach in an integrated multicategory purchase timing model. Managers could then benefit from the potential insights from such an approach by deriving customized marketing efforts across categories.

In this paper, our main focus has been on the accurate individual-level prediction of the expected number of future transactions, which could easily be converted into a discounted quantity to yield a net present value as suggested by Fader et al. (2005b). Although (discounted) expected transactions are an important aspect of customer valuation, extending the Pareto/GGG toward a full CLV model would require a submodel

for the purchase amount per transaction. The flexibility of our hierarchical Bayesian model approach permits the incorporation of such an extension, for example, by assuming a standard normal (Schmittlein and Peterson 1994), a log-normal (Borle et al. 2008), or a gamma-gamma (Colombo and Jiang 1999) submodel for purchase amounts. With such an extension toward a fully faceted model to predict residual CLV for a customer base, relationship managers could benefit further from the insights we have gained from applying the Pareto/GGG in our empirical studies.

To our knowledge, this research represents the first systematic study demonstrating that the presence of ITT regularities improves the predictability of future purchase behavior. Regarding further research, we anticipate similar gains from our proposed gamma-distributed timing model for inferring customers' latent activity states not only in BTYD settings but also for the broader class of hidden Markov models (Schwartz et al. 2014),<sup>10</sup> with the promise of detecting changes between high-frequency and low-frequency purchase phases (or vice versa) more quickly. This also includes models in which customers are allowed to make back-and-forth transitions between an active and an inactive state (e.g., an "on and off" purchasing model; see Schwartz et al. 2014). We conjecture that modeling approaches accounting for such nonstationary repurchase behavior might be good candidates for capturing clumpy ITT patterns (interpreted as "episodes" with higher purchase propensities followed by a period with lower or even no activity) and to translate this capability into better predictions of future transactions. However, such models typically come at some additional costs because they require complete purchase histories and not merely summary statistics.

Certainly, we also have to acknowledge that, similar to any other BTYD model, the Pareto/GGG implicitly assumes stationary marketing activities in both the calibration and forecasting periods. It thus can serve as a baseline for benchmarking the impact of changes in target marketing actions (Fader et al. 2005a). It is beyond the scope of this paper but would be important yet challenging to build a model that incorporates marketing covariates, in addition to accounting for ITT regularities. The hidden Markov model-based approaches presented by Netzer et al. (2008) and Schweidel et al. (2011) offer promising starting points for endeavors to model the interplay between ITT regularity and marketing actions.

Finally, several other extensions of our research would be welcome. In particular, we call for studies that translate the general idea underlying the Pareto/GGG

<sup>10</sup> Note that BTYD models are also constrained variants of hidden Markov models, with two states, one of which is an absorbing, inactive state.

into a discrete-time setting. The BG/BB introduced by Fader et al. (2010) could serve as a modeling framework, although researchers would need to relax the memoryless binomial assumption for purchase occurrences. In addition, the modeling flexibility gained by an MCMC sampling scheme might facilitate links between individual-level parameters ( $\lambda$ ,  $\mu$ , but also  $k$ ) and customer-specific covariates, as well as allow for correlations between them, as demonstrated by Abe (2009). With a link between the purchase and dropout processes, the effect of ITT regularities on customers' activity states could be studied even more thoroughly than we have done in this study.

### Supplemental Material

Supplemental material to this paper is available at <http://dx.doi.org/10.1287/mksc.2016.0963>.

### Acknowledgments

The authors thank the editor-in-chief, the associate editor, and the anonymous reviewers for their helpful suggestions and valuable guidance throughout the review process. This research also benefited from comments and input by Sandeep Conoor, Peter Fader, Nicolas Gladly, Bruce Hardie, Daniel McCarthy, Udo Wagner, and the audience at the Marketing Science Conference 2014. The authors also thank Eric Bradlow and Yao Zhang for sharing their data sets.

### Appendix A. Derivation of $P(\text{alive})$

We provide the derivation of the probability of a customer still being alive at time  $T$  here (Schmittlein et al. 1987, Appendix 1). Let  $\theta$  denote the individual-level parameters  $\{k, \lambda, \mu\}$ , let  $\xi$  indicate the observed data  $\{t_1, \dots, t_x, T\}$ , and let  $\phi_{T-t_x}$  refer to the event of no transaction occurring in  $(t_x, T]$ . Then,

$$P(\tau > T | \theta, \xi) = \frac{f_1(\xi | \theta, \tau > T)P(\tau > T | \theta)}{f_1(\xi | \theta, \tau > T)P(\tau > T | \theta) + f_2(\xi, t_x < \tau \leq T | \theta)}. \quad (\text{A1})$$

The assumption of exponentially distributed lifetimes gives us

$$P(\tau > T | \theta) = e^{-\mu T}. \quad (\text{A2})$$

The likelihood functions  $f_1$  and  $f_2$  can be further split into independent components

$$f_1(\xi | \theta, \tau > T) = f_3(\xi | t_x, \phi_{T-t_x}, \theta, \tau > T) \cdot f_4(t_x, \phi_{T-t_x} | \theta, \tau > T) \quad \text{and} \quad (\text{A3})$$

$$f_2(\xi, t_x < \tau \leq T | \theta) = f_5(\xi | t_x, \phi_{\tau-t_x}, \theta) \cdot f_6(t_x, \phi_{\tau-t_x}, t_x < \tau \leq T | \theta). \quad (\text{A4})$$

Conditional on the time of the last transaction  $t_x$ , the exact timing of the earlier transactions  $t_1, \dots, t_{x-1}$  is independent of the subsequent timing of the dropout. Therefore,  $f_3 = f_5$ , and the according terms cancel out in Equation (A1).

Conditional on the dropout  $\tau > T$ , the timing of the last transaction  $t_x$  and an observed waiting time of  $T - t_x$  are independent events that can be derived separately. Because the sum of  $x$  independent and identically distributed variables

$\Delta t_j \sim \text{Gamma}(k, k\lambda)$  is a gamma-distributed random variable with an updated shape parameter  $kx$ , it follows that

$$\begin{aligned} f_4(t_x, \phi_{T-t_x} | \theta, \tau > T) &= f(t_x | x, \theta) f(\phi_{T-t_x} | \theta) \\ &= \frac{(k\lambda)^{kx}}{\Gamma(kx)} t_x^{kx-1} e^{-k\lambda t_x} \\ &\quad \cdot (1 - F_T(T - t_x | k, k\lambda)). \end{aligned} \quad (\text{A5})$$

Similarly,  $f_6$  can be expressed by integrating  $f_4$  over  $t_x < \tau \leq T$ , such that  $\tau$  is exponentially distributed

$$\begin{aligned} f_6(t_x, \phi_{\tau-t_x}, t_x < \tau \leq T | \theta) &= \int_{t_x}^T f_4(t_x, \phi_{y-t_x} | \theta, y > t_x) f(\tau = y | \theta) dy \\ &= \frac{(k\lambda)^{kx}}{\Gamma(kx)} t_x^{kx-1} e^{-k\lambda t_x} \int_{t_x}^T (1 - F_T(y - t_x | k, k\lambda)) \mu e^{-\mu y} dy. \end{aligned} \quad (\text{A6})$$

Putting it all together, we obtain

$$\begin{aligned} P(\tau > T | k, \lambda, \mu, t_1, \dots, t_x, T) &= \left( 1 + \frac{\int_{t_x}^T (1 - F_T(y - t_x | k, k\lambda)) \mu e^{-\mu y} dy}{(1 - F_T(T - t_x | k, k\lambda)) e^{-\mu T}} \right)^{-1}. \end{aligned} \quad (\text{A7})$$

For the degenerate case of  $k = 1$ , this expression can be simplified to the Pareto/NBD result published by Schmittlein et al. (1987)

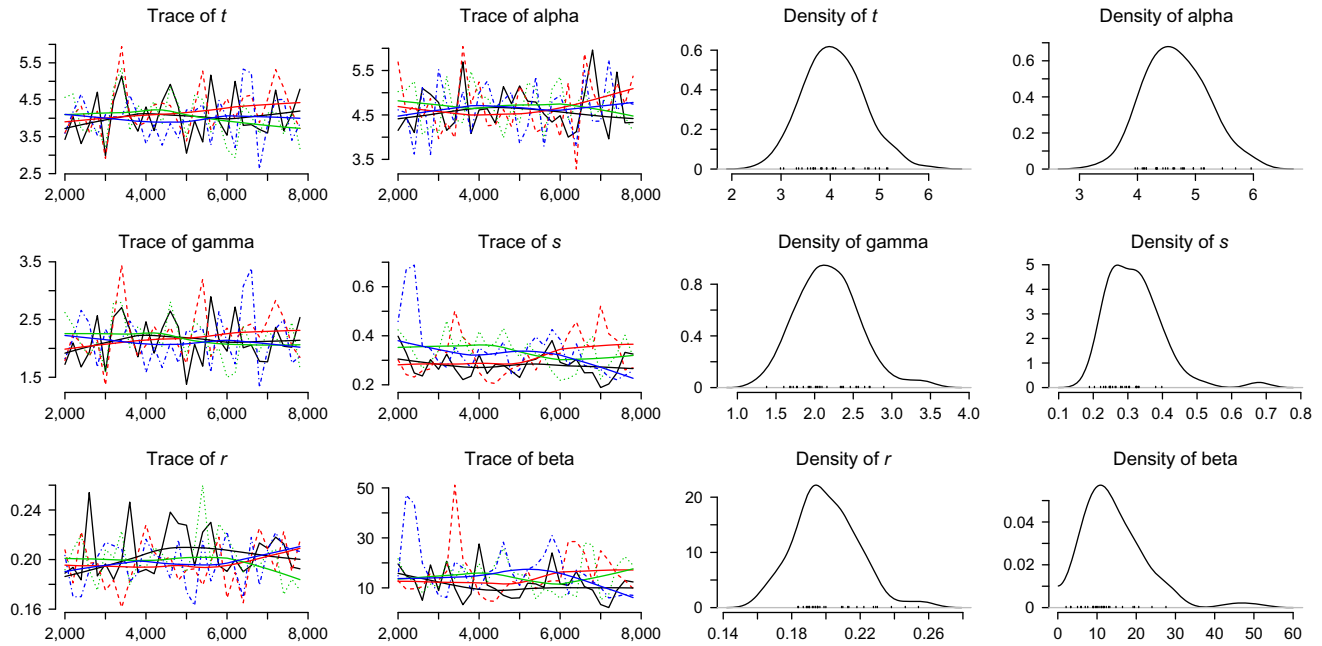
$$\begin{aligned} F_T(T - t_x | 1, \lambda) &= 1 - e^{-\lambda(T-t_x)} \\ P(\text{alive}) &= \left( 1 + \frac{\int_{t_x}^T e^{-\lambda(y-t_x)} \mu e^{-\mu y} dy}{e^{-\lambda(T-t_x)} e^{-\mu T}} \right)^{-1} \\ &= \left( 1 - \frac{\mu}{\lambda + \mu} \frac{e^{-y(\lambda+\mu)} \Big|_{t_x}^T}{e^{-(\lambda+\mu)T}} \right)^{-1} \\ &= \left( 1 - \frac{\mu}{\lambda + \mu} (1 - e^{-(T-t_x)(\lambda+\mu)}) \right)^{-1}. \end{aligned}$$

### Appendix B. Further Details on Simulation Study

In total, 160 data sets based on a variety of parameter settings for  $N$ ,  $r$ ,  $\alpha$ ,  $s$ ,  $\beta$ ,  $t$ , and  $\gamma$  have been generated and used for assessing the predictive accuracy of the Pareto/GGG and Pareto/NBD models. For each setting, we ran four separate MCMC chains with 8,000 iterations each, then retained only every 200th iteration after an initial burn-in of 2,000 iterations. To check for convergence, we used the Gelman diagnostic (Gelman and Rubin 1992). Figure B.1 depicts an example MCMC run, with the left-hand side showing the trace plots of four separate MCMC chains for each of the six heterogeneity parameters, and the right-hand side the corresponding sampled posterior densities. Running this configuration for 1,000 customers requires 160 million individual-level parameter draws (1,000 customers  $\times$  5 parameters  $\times$  4 chains  $\times$  8,000 iterations); for our R/Rcpp (R Core Team 2014, Eddelbuettel and François 2011) implementation, on a laptop equipped with a 2.5 GHz quad-core Intel Core i7 chip, running the four chains in parallel took approximately four minutes. A GPU-based implementation could speed up the runtime even further (White and Porter 2014).

Tables B.1 and B.2 report the relative as well as absolute lift in customer-level mean absolute error for the holdout period for all 160 simulated worlds.

**Figure B.1** (Color online) MCMC Chains for the Simulated Scenario with  $t = 5$ ,  $\gamma = 2.5$ ,  $r = 0.25$ ,  $\alpha = 5$ ,  $s = 0.25$ ,  $\beta = 5$ , and  $N = 1,000$



**Table B.1** Relative Lift in Customer-Level MAE for Holdout Period

			$N = 1,000$ (%)				$N = 4,000$ (%)				Mean (%)
	$s$	$r$ $\beta \backslash \alpha$	0.25 5	0.75 5	0.25 15	0.75 15	0.25 5	0.75 5	0.25 15	0.75 15	
$t = 1.6/y = 0.4$ $\text{mean}(k) = 4$	0.25	5	+8	+6	+5	+12	+8	+11	+4	+11	+8
	0.75	5	+11	+10	+11	+19	+11	+10	+22	+13	+14
	0.25	15	+10	+10	+12	+12	+6	+8	+8	+10	+10
	0.75	15	+16	+10	+20	+17	+11	+11	+12	+16	+14
$t = 5/y = 2.5$ $\text{mean}(k) = 2$	0.25	5	+5	+5	+6	+9	+5	+6	+6	+8	+6
	0.75	5	+4	+4	+17	+12	+7	+7	+19	+10	+10
	0.25	15	+4	+7	+5	+11	+4	+5	+7	+7	+6
	0.75	15	+7	+4	+14	+11	+7	+6	+14	+9	+9
$t = 6/y = 4$ $\text{mean}(k) = 1.5$	0.25	5	+3	+4	0	+3	+3	+2	+4	+2	+3
	0.75	5	+7	+5	+13	+7	+3	+3	+10	+5	+7
	0.25	15	+4	+4	+4	+5	+2	+2	+3	+3	+3
	0.75	15	+6	0	+7	+9	+5	+3	+10	+5	+5
$t = 8/y = 8$ $\text{mean}(k) = 1$	0.25	5	-4	0	0	0	0	-2	0	-2	-1
	0.75	5	+6	-6	+14	0	+5	0	+11	+3	+4
	0.25	15	-1	0	0	-2	-2	0	-2	-2	-1
	0.75	15	0	0	+9	-3	0	0	+2	0	0
$t = 17/y = 20$ $\text{mean}(k) = 0.85$	0.25	5	0	-1	0	-2	-1	-2	-2	-2	-1
	0.75	5	+1	0	+7	+2	-2	-3	+4	0	+1
	0.25	15	-1	-3	-2	-4	-1	-2	-3	-2	-2
	0.75	15	-3	-1	+1	-4	-1	-2	+1	-3	-2
Mean			+4	+3	+7	+6	+3	+3	+7	+4	

**Table B.2** Absolute Lift in Customer-Level MAE for Holdout Period

			$N = 1,000$				$N = 4,000$				Mean
	$s$	$r$ $\beta \backslash \alpha$	0.25 5	0.75 5	0.25 15	0.75 15	0.25 5	0.75 5	0.25 15	0.75 15	
$t = 1.6/y = 0.4$ $\text{mean}(k) = 4$	0.25	5	+0.05	+0.09	+0.02	+0.11	+0.06	+0.18	+0.02	+0.10	+0.08
	0.75	5	+0.03	+0.06	+0.01	+0.06	+0.03	+0.07	+0.03	+0.04	+0.04
	0.25	15	+0.08	+0.19	+0.05	+0.12	+0.05	+0.16	+0.04	+0.10	+0.10
	0.75	15	+0.08	+0.13	+0.05	+0.11	+0.06	+0.14	+0.03	+0.10	+0.09



Table B.2 (Continued)

	<i>s</i>	<i>r</i> $\beta \backslash \alpha$	<i>N</i> = 1,000				<i>N</i> = 4,000				Mean
			0.25	0.75	0.25	0.75	0.25	0.75	0.25	0.75	
			5	5	15	15	5	5	15	15	
$t = 5/v = 2.5$	0.25	5	+0.04	+0.08	+0.02	+0.08	+0.04	+0.10	+0.02	+0.08	+0.06
mean( <i>k</i> ) = 2	0.75	5	+0.01	+0.02	+0.03	+0.04	+0.02	+0.05	+0.02	+0.03	+0.03
	0.25	15	+0.03	+0.13	+0.02	+0.12	+0.04	+0.10	+0.03	+0.08	+0.07
	0.75	15	+0.03	+0.05	+0.04	+0.07	+0.03	+0.07	+0.04	+0.06	+0.05
$t = 6/y = 4$	0.25	5	+0.02	+0.06	+0.00	+0.03	+0.02	+0.04	+0.02	+0.02	+0.03
mean( <i>k</i> ) = 1.5	0.75	5	+0.02	+0.03	+0.02	+0.02	+0.01	+0.02	+0.01	+0.02	+0.02
	0.25	15	+0.03	+0.08	+0.02	+0.05	+0.02	+0.05	+0.02	+0.04	+0.04
	0.75	15	+0.03	+0.00	+0.02	+0.06	+0.03	+0.04	+0.03	+0.03	+0.03
$t = 8/y = 8$	0.25	5	−0.03	−0.01	−0.00	−0.00	−0.00	−0.04	−0.00	−0.02	−0.01
mean( <i>k</i> ) = 1	0.75	5	+0.02	−0.04	+0.02	+0.00	+0.01	+0.01	+0.02	+0.01	+0.01
	0.25	15	−0.01	−0.00	−0.00	−0.02	−0.02	−0.02	−0.01	−0.02	−0.01
	0.75	15	+0.00	−0.01	+0.03	−0.02	−0.00	−0.01	+0.01	−0.00	−0.00
$t = 17/y = 20$	0.25	5	−0.00	−0.02	−0.00	−0.02	−0.01	−0.03	−0.01	−0.03	−0.02
mean( <i>k</i> ) = 0.85	0.75	5	+0.01	−0.00	+0.01	+0.01	−0.01	−0.02	+0.01	−0.00	+0.00
	0.25	15	−0.01	−0.07	−0.01	−0.05	−0.01	−0.05	−0.01	−0.03	−0.03
	0.75	15	−0.02	−0.02	+0.00	−0.03	−0.01	−0.03	+0.00	−0.02	−0.01
Mean			+0.02	+0.04	+0.02	+0.04	+0.02	+0.04	+0.01	+0.03	

## References

- Abe M (2009) Counting your customers one by one: A hierarchical Bayes extension to the Pareto/NBD model. *Marketing Sci.* 28(3):541–553.
- Allenby GM, Leone RP, Jen L (1999) A dynamic model of purchase timing with application to direct marketing. *J. Amer. Statist. Assoc.* 94(446):365–374.
- Barabasi A-L (2005) The origin of bursts and heavy tails in human dynamics. *Nature* 435:207–211.
- Bardhan I, Oh JH(C), Zheng Z(E), Kirksey K (2015) Predictive analytics for readmission of patients with congestive heart failure. *Inform. Systems Res.* 26(1):19–39.
- Batistam EP, Denizel M, Filiztekin A (2007) Empirical validation and comparison of models for customer base analysis. *Internat. J. Res. Marketing* 24(3):201–209.
- Bemmar AC, Gladys N (2012) Modeling purchasing behavior with sudden death: A flexible customer lifetime model. *Management Sci.* 58(5):1012–1021.
- Borle S, Singh SS, Jain DC (2008) Customer lifetime value measurement. *Management Sci.* 54(1):100–112.
- Chatfield C, Goodhardt GJ (1973) A consumer purchasing model with Erlang inter-purchase time. *J. Amer. Statist. Assoc.* 68(344):828–835.
- Colombo R, Jiang W (1999) A stochastic RFM model. *J. Interactive Marketing* 13(3):2–12.
- Conoor SS (2010) Customer-base analysis in noncontractual settings. Unpublished doctoral thesis, Northwestern University, Evanston, IL.
- Dunn R, Reader S, Wrigley N (1983) An investigation of the assumptions of the NBD model as applied to purchasing at individual stores. *Appl. Statist.* 32(3):249–259.
- Dziurzynski L, Wadsworth E, McCarthy D (2014) BTYD: Implementing buy 'til you die models. <http://CRAN.R-project.org/package=BTYD>.
- Eddelbuettel D, François R (2011) Rcpp: Seamless R and C++ integration. *J. Statist. Software* 40(8):1–18.
- Fader PS (2013) *Customer Centricity: Focus on the Right Customers for Strategic Advantage* (Wharton Executive Essentials, Philadelphia).
- Fader PS, Hardie BGS (2001) Forecasting repeat sales at CDNOW: A case study. *Interfaces* 31(3 suppl):S94–S107.
- Fader PS, Hardie BGS (2007) Incorporating time-invariant covariates into the Pareto/NBD and BG/NBD models. <http://www.brucehardie.com/notes/019>.
- Fader PS, Hardie BGS (2009) Probability models for customer-base analysis. *J. Interactive Marketing* 23(1):61–69.
- Fader PS, Hardie BGS, Chun-Yao H (2004) A dynamic changepoint model for new product sales forecasting. *Marketing Sci.* 23(1):50–65.
- Fader PS, Hardie BGS, Lee KL (2005a) "Counting your customers" the easy way: An alternative to the Pareto/NBD model. *Marketing Sci.* 24(2):275–284.
- Fader PS, Hardie BGS, Lee KL (2005b) RFM and CLV: Using ISO-value curves for customer base analysis. *J. Marketing Res.* 42(4):415–430.
- Fader PS, Hardie BGS, Shang J (2010) Customer-base analysis in a discrete-time noncontractual setting. *Marketing Sci.* 29(6):1086–1108.
- Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Statist. Sci.* 7(5):457–472.
- Gupta S (1991) Stochastic models of interpurchase time with time-dependent covariates. *J. Marketing Res.* 28(1):1–15.
- Gupta S, Hanssens D, Hardie BGS, Kahn W, Kumar V, Lin N, Ravishanker N, Sriram S (2006) Modeling customer lifetime value. *J. Service Res.* 9(2):139–155.
- Herniter J (1971) A probabilistic market model of purchase timing and brand selection. *Management Sci.* 18(4):102–112.
- Hoppe D, Wagner U (2007) Customer base analysis: The case for a central variant of the Betageometric/NBD model. *Marketing ZFP-J. Res. Management* 3(2):75–90.
- Jerath K, Fader PS, Hardie BGS (2011) New perspectives on customer "death" using a generalization of the Pareto/NBD model. *Marketing Sci.* 30(5):866–880.
- Kumar V (2008) *Customer Lifetime Value: The Path to Profitability* (Now Publishers, Hanover, MA).
- Kumar V, Venkatesan R, Bohling T, Beckmann D (2008) The power of CLV: Managing customer lifetime value at IBM. *Marketing Sci.* 27(4):585–599.
- Ma SH, Liu JL (2007) The MCMC approach for solving the Pareto/NBD model and possible extensions. *Proc. Third Internat. Conf. Natural Comput.*, Vol. 2 (IEEE Computer Society, Los Alamitos, CA), 505–512.
- Malthouse EC (2001) Assessing the performance of direct marketing scoring models. *J. Interactive Marketing* 15(1):49–62.
- Malthouse EC (2009) The results from the lifetime value and customer equity modeling competition. *J. Interactive Marketing* 23(3):272–275.

- Malthouse EC, Blattberg RC (2005) Can we predict customer lifetime value? *J. Interactive Marketing* 19(1):2–16.
- Morrison DG, Schmittlein DC (1988) Generalizing the NBD model for customer purchases: What are the implications and is it worth the effort? *J. Bus. Econom. Statist.* 6(2):145–159.
- Neal RM (2003) Slice sampling. *Ann. Statist.* 31(3):705–741.
- Netzer O, Lattin JM, Srinivasan V (2008) A hidden Markov model of customer relationship dynamics. *Marketing Sci.* 27(2):185–204.
- Quintana FA, Marshall P (2015) A Bayesian non-parametric Pareto/NBD model: Individual and cluster analysis. Working paper, Pontificia Universidad Católica de Chile, Santiago, Chile.
- R Core Team (2014) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna).
- Reinartz WJ, Kumar V (2000) On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing. *J. Marketing* 64(4):17–35.
- Rossi PE, Allenby GM (2003) Bayesian statistics and marketing. *Marketing Sci.* 22(3):304–328.
- Schmittlein DC, Morrison DG (1983) Prediction of future random events with the condensed negative binomial distribution. *Internat. Amer. Statist. Assoc.* 78(382):449–456.
- Schmittlein DC, Peterson RA (1994) Customer base analysis: An industrial purchase process application. *Marketing Sci.* 13(1):41–67.
- Schmittlein DC, Morrison DG, Colombo R (1987) Counting your customers: Who are they and what will they do next? *Management Sci.* 33(1):1–24.
- Schwartz EM, Bradlow ET, Fader PS (2014) Model selection using database characteristics: Developing a classification tree for longitudinal incidence data. *Marketing Sci.* 33(2):188–205.
- Schweidel DA, Fader PS (2009) Dynamic changepoints revisited: An evolving process model of new product sales. *Internat. J. Res. Marketing* 26(2):119–124.
- Schweidel DA, Knox G (2013) Incorporating direct marketing activity into latent attrition models. *Marketing Sci.* 32(3):471–487.
- Schweidel DA, Bradlow ET, Fader PS (2011) Portfolio dynamics for customers of a multiservice provider. *Management Sci.* 57(3):471–486.
- Shah D, Rust R, Parasuraman A, Staelin R, Day GS (2006) The path to customer centricity. *J. Service Res.* 9(2):113–124.
- Singh SS, Borle S, Jain DC (2009) A generalized framework for estimating customer lifetime value when customer lifetimes are not observed. *Quant. Marketing Econom.* 7(2):181–205.
- Tanner MA, Wong WH (1987) The calculation of posterior distributions by data augmentation. *J. Amer. Statist. Assoc.* 82(398):528–540.
- Tirenni G, Labbi A, Berrospi C, Elisseeff A, Bhose T, Pauro K, Pöyhönen S (2007) Customer equity and lifetime management (CELM) Finnair case study. *Marketing Sci.* 26(4):553–565.
- Wheat RD, Morrison DG (1990) Estimating purchase regularity with two interpurchase times. *J. Marketing Res.* 27(1):87–93.
- White G, Porter MD (2014) GPU accelerated MCMC for modeling terrorist activity. *Comput. Statist. Data Anal.* 71:643–651.
- Winer RS (2001) A framework for customer relationship management. *California Management Rev.* 43(4):89–105.
- Wu C, Chen H-L (2000) A consumer purchasing model with learning and departure behaviour. *J. Oper. Res. Soc.* 51(5):583–591.
- Wübben M, von Wangenheim F (2008) Instant customer base analysis: Managerial heuristics often “get it right.” *J. Marketing* 72(3):82–93.
- Zhang Y, Bradlow ET, Small DS (2013) New measures of clumpiness for incidence data. *J. Appl. Statist.* 40(11):2533–2548.
- Zhang Y, Bradlow ET, Small DS (2015) Predicting customer value using clumpiness: From RFM to RFMC. *Marketing Sci.* 34(2):195–208.