Customer base analysis frequently relies on models to describe and predict long-term behavior of customers. The “gold standard” for the analysis of non-contractual settings is the well-known Pareto/NBD model. The Betageometric/NBD model was proposed as an easy-to-calibrate drop-in replacement for the latter. This article demonstrates that there are customer bases where the fit of the Betageometric/NBD might be improved. This is accomplished by relaxing the assumption regarding the defection behavior of customers, thereby increasing the versatility of the approach. The resulting model, the central Betageometric/NBD model, provides the same statistics and possesses the same number of parameters, but produces a better fit for certain cases, i.e., when a substantial proportion of customers drop out after trial. This is demonstrated by applying the model to two datasets from a catalog retailer, where the central Betageometric/NBD dominates both the Pareto/NBD and the Betageometric/NBD.

Keywords
consumer behavior, purchase behavior, trial and repeat buying, probability models, customer lifetime, catalog retailer

1. Introduction
When planning marketing activities, managers typically face the problem of targeting the customers who have the highest likelihood of responding positively to a campaign by purchasing the goods or services offered. In direct marketing and other businesses where the relationship to the customer is not based on contractual agreements, the class of “Recency, Frequency and Monetary Value” (RFM) models serves as a means of identifying clients likely to respond to a marketing stimulus, both in practitioners’ literature, e.g., Weinstein (2004, p. 85) and in scholarly journals, e.g., Bitran/Mondschein (1996). RFM models rely on three pieces of information to assign customers to RFM-segments: recency (the time since the last purchase), frequency (the number of purchases observed so far), and monetary value (usually defined as the cumulative sales volume). Test stimuli are then sent to a small sub-sample of individuals in each such segment in order to obtain likely response rates for a full-scale campaign. Marketers finally restrict their campaign to those segments where the expected revenue meets the specified goals.

While RFM models are easy to use, they lack a solid theoretical foundation. In practice, these models are often simplified to “RF” or even to elementary rules based on recency only. In the latter situation, customers are considered to have quit business if their last purchase occurred prior to a predefined distance in time. It is obvious that such an approach is not optimal if customers differ in their purchase rates – and the evidence for heterogeneous purchase rates is overwhelming. Therefore, different authors have followed up on this critique and developed a class of integrated stochastic purchase-incidence/lifetime-duration models.

Schmittlein/Morrison/Colombo’s seminal paper (1987) concentrates on customers’ purchase intensities and customers’ lifetimes [1]. The observed number of purchases within a certain time period provides a clear indication for the first aspect. However, there is usually no apparent information for the latter available since, typically, clients quit business without giving explicit notice; a long time period having elapsed since his/her last purchase is only a partial indicator of defection. SMC overcome this problem by postulating rather simple assumptions about the individual lifetime processes which makes it possible to infer whether a customer is likely to be still active or not. In the very spirit of stochastic models of buying behavior, SMC build upon empirically observed regulari-
ties and assume that purchase frequencies may be adequately described by a negative binomial (NBD) model and customers’ lifetimes by a Pareto model. They justify these assumptions in detail and provide mathematical expression in order to address marketing problems such as:

- How many customers does the firm presently have?
- How has the customer base grown over the past years?
- Which clients most likely represent active customers?
- What level of transactions should be expected in the future by those existing clients?

SM C’s paper earned great respect because their Pareto/NBD has several desirable properties. However, the model was found challenging to implement. The estimation of parameters and the computation of model measures were identified as complex tasks. These obstacles are reflected in a mismatch between references in the literature (numerous) and empirical applications (few). Fader/ Hardie/Lee (2005a) probe this issue and develop a variation of SM C’s Pareto/NBD, the Betageometric/NBD [2]. This model is an elegant and powerful alternative since it is much easier to implement (e.g., just by using Microsoft Excel) and yields very similar results in a wide variety of purchasing environments. In particular, rather than assuming exponentially distributed lifetimes on the individual level, FHL assume that customers may become inactive after any transaction according to a shifted geometric distribution. Since FHL use a customer’s first purchase to decide whether this client belongs to the cohort which should be analyzed (and exclude this observation from subsequent calculations – cf. Section 2.2), a customer may quit business no earlier than after his/her first repeat purchase. This assumption might be unrealistic in situations where a significant portion of consumers drop out after their first trial or, as Parfitt/Collins (1968) put it: “Getting [...] to try the brand is comparatively easy. [...] The ultimate success of any brand depends on the willingness of consumers, once having tried it, to continue purchasing it.” Consequently, Parfitt/Collins (1968) model trial and repeat separately.

The paper at hand tries to overcome this limitation by relaxing FHL’s assumption and allowing customers to drop out immediately after trial, according to a central geometric distribution. All other assumptions remain unchanged. From a managerial point of view, the results provided address the same issues as the work by SM C and FHL, i.e., who are the active clients in the customer base? What is the predicted level of business generated by certain segments of them? From a methodological point of view, the derivation of the model is carried out from a new perspective, focusing on count processes (FHL tackle this task by means of a timing process). This makes it possible to reduce mathematical sophistication. In addition, it will be shown that a restriction on the parameter range for FHL’s model and our variant may be relaxed since the functional discontinuity causing this restriction may be removed. Finally, from a computational point of view the model follows FHL and may also be implemented in commercially available spreadsheet software.

The remainder of the paper is organized as follows. Section 2 describes the proposed approach – a central Betageometric/NBD – from a conceptual perspective and provides details on how to calculate measures of managerial relevance. A replication study is presented in Section 3.1, and it reveals that the proposed model slightly outperforms FHL’s model based on the same dataset that FHL analyze in their paper. Section 3.2 compares the Pareto/NBD (P/NBD), the Betageometric/NBD (BG/NBD) and the proposed central variant of the Betageometric/NBD (CBG/NBD) with respect to fit and predictive accuracy for empirical data of a medium sized, German catalog retailer and confirms that the latter performs very well. A summary (Section 4) concludes this article.

2. The Central Variant of the Betageometric/NBD

2.1. Data Requirements

All of the models introduced above rely on individual level data on the number and the timing of the customers’ previous transactions. In particular, the following notation will be used:

\[ x \] is the number of purchases [3] a customer [4] made in a given time span (excluding a possible purchase at time zero).

\[ t_x \] is the time of his/her last purchase, i.e., the time of purchase incident \( t_x \) in “model time” (explained below). \( t_x \) is required to be strictly positive, but for notational convenience \( t_x := 0 \) if no purchases are observed at all for this customer.

\( T \) is the time of reference of the customer in model time.

These data requirements are closely related to those for RFM-analyses. Nevertheless, the meaning of model time and the restriction to purchases in the time interval \((0, T]\) requires further explanation. A difference between calendar and model time occurs because, usually, customers are pooled into cohorts in order to end up with a sample of sufficient size. These cohorts are formed according to clients experiencing a common event within the same time interval. It is routine practice to take a patron’s first purchase as this common event but other definitions might be applied as well; by definition, this common event takes place at customer’s time zero. If the analysis is carried out for a fixed day (in calendar time) each customer’s preceding time of reference \( T \) will vary depending on the specific day of his/her first purchase within the pooling interval. The same considerations apply for a customer’s time zero and the time of his/her last purchase \( t_x \) [5]. The exclusion of the time zero appears to be a restriction but leads to increased adaptability, i.e., the model’s user can flexibly choose the event that constitutes a customer acquisition for the market under consideration. This might be an initial order in one scenario, a
2.2. Behavioral Assumptions

As outlined above, the only difference between FHL’s model and its variant proposed here is the way in which events occurring immediately at time zero are handled. The postulated behavioral assumptions are almost identical.

(A1) Individual purchases of customers follow a Poisson process with rate $\lambda$.

(A2) The individual level purchase rates $\lambda$ are distributed gamma across customers with shape parameter $r$ and scale parameter $\alpha$.

(A3) Customers have a limited lifetime. At time zero and directly after each purchase there is a constant probability $p$ ($0 \leq p \leq 1$) that the customer will become inactive. Therefore, dropouts follow a central geometric distribution with dropout probability $p$.

(A4) The individual level dropout probabilities $p$ are distributed beta across customers with the shape parameters $a$ and $b$.

(A5) Purchase process and lifetime process are independent of each other.

Although neither SMC nor FHL state it explicitly, it is also assumed that market conditions are sufficiently stable to warrant stationarity. Since the justification of these assumptions has already been provided by SMC and FHL, we concentrate on our modification in (A3), i.e., to additionally allow for defection at time zero. In order to motivate this change, the acquisition of new customers is discussed through the examples of fictive companies:

Company A is an Internet store selling, e.g., books, journals, or apparel. Typically, when a new customer of company A experiences a need for A’s products or services, either impulsive as a response to marketing stimuli, or planned, he/she signs up on the Web site and places his/her initial order. Here, the arrival of a new customer coincides with the first order. He/she can now either stay with the business and place subsequent orders or drop out for a variety of reasons.

Company B is a classic catalog retailer. It also has some clients that sign up of their own accord for their services. In direct marketing campaigns, B additionally makes regular use of rental lists – external databases of prospective customers. Some customers will place initial and subsequent orders. Many customers from these rental lists, however, will not place an initial order with B at all.

Company C is a cellular phone service provider with a prepaid cellular phone product. The odds of receiving at least some recharges are fairly high for C. Their clients have made a substantial initial investment into a cellular phone and this device will, in most cases, be limited for use exclusively in company C’s network. Furthermore, customers may experience a considerable switching cost when informing family, friends, and business partners about a changed telephone number (Shi/Chiang/Rhee 2006 discuss such switching cost in detail).

As outlined above, the P/NBD and the BG/NBD have one important property in common: both models do not account for time zero formally, put $t_0 \in (0,T]$. Notwithstanding, they differ with respect to the timing of a potential dropout. The P/NBD postulates exponentially distributed lifetimes and therefore customers may defect at any time greater than zero. In contrast, the BG/NBD replaces the continuous lifetimes of the P/NBD with a discrete geometric process where clients can defect directly after purchases only. The assumption seems reasonable that the propensity to defect is particularly high immediately after a purchase. This might be the case, e.g., because of dissatisfaction with the consumed good or service (in reality a customer might decide to leave at any point in time or directly before purchasing at another company).

The peculiarity of the BG/NBD does not inherently arise with the discrete process, but rather with the interval covered, as visualized in Figure 1. The initial purchase at company A’s Web site is not considered by either model – customers cannot defect at time zero. Still, the P/NBD allows defection at any time $0 + \varepsilon$ for arbitrarily small values of $\varepsilon$ without a preceding purchase. Contrarily, the design of the BG/NBD does not allow a customer to defect prior to placing an order in $(0,T)$, hence not before $t_0$. Those customers who did not purchase within the whole observation horizon (i.e., $x = 0$, $t_0 = 0$) are considered active by the BG/NBD by definition. This makes the model assumptions questionable in at least two of the three scenarios outlined above. Company A, the Internet retailer,
and company B, the direct marketing company using rental lists, will usually experience a high proportion of never-rebuyers (company A) or non-buyers (company a). Only for company c is this assumption thoroughly plausible for the reasons outlined earlier.

This design feature restricts the flexibility of the BG/NBD and impacts its ability to adapt to datasets with few repeat-after-trial purchases. We show that this restriction can easily be removed by incorporating an initial opportunity to defect at time zero (cf. A3) without changing the overall structure of the model. Attrition therefore follows a central geometric distribution, because the geometric process starts at time zero and constitutes the central BG/NBD model. The difference between the (shifted) original model and our variant is visualized by the CBG/NBD line in Figure 1: the CBG/NBD still covers purchase incidences in the interval [0,T], but an additional dropout opportunity at time zero is introduced, so that purchase and possible dropout at t are both conditional on the customer not having defected at time zero, symbolized by the dashed line.

### 2.3. Key Expressions and Model Results

The CBG/NBD belongs to the class of integrated stochastic purchase-incidence/lifetime-duration models. Although no standardized notation is used in the literature, we closely follow FHL to facilitate comparisons; the symbols used are summarized in Table 1. In the same vein, we will consider a similar set of measures for customer base analysis. These measures provide important managerial insights; some examples are listed in Table 2. In this subsection, we will explicitly derive three of these expressions: the likelihood function, the probability distribution of purchase frequencies, and the probability of a customer still being active.

The derivation will be carried out by means of count processes which allow an intuitive interpretation of several of the mathematical expressions. It can be shown, however, that a derivation which closely follows FHL yields identical results. In order to increase the readability of this paper, the mathematical details are kept at a minimal level. Rigorous derivations of all expressions can be found in a supplementary appendix that is available upon request from the authors.

#### 2.3.1. Likelihood Function

Postulating (A1)–(A5) we derive the likelihood function as follows. While an individual customer is active, his/her purchases follow a Poisson distribution (A1). With x purchases in the interval from 0 to t, the customer cannot have defected on the previous x dropout opportunities (after time 0, t, t, ..., t). Thus, the probability of observing these x purchases is the Poisson probability for an interval of length t. In the time interval [t, T], no further purchases occurred. Consequently, the client either defected directly after t, or he/she did not rebuy by time T while still being active, thus \( P(X(T - t) = 0) = p + (1 - p)P_{\text{Pois}}(X(T - t) = 0) \). Taking into account the

### Table 1: Symbols used

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t )</td>
<td>time</td>
</tr>
<tr>
<td>( x )</td>
<td>number of transactions/purchases of a customer</td>
</tr>
<tr>
<td>( t_x )</td>
<td>time of the last observed transaction of a customer</td>
</tr>
<tr>
<td>( T )</td>
<td>time of reference of a customer, typically the end of the observation horizon</td>
</tr>
<tr>
<td>( \lambda, \ p )</td>
<td>individual level purchase rate and defection probability</td>
</tr>
<tr>
<td>( r, \alpha )</td>
<td>shape and scale parameter of the NBD purchase process</td>
</tr>
<tr>
<td>( a, b )</td>
<td>shape parameters of the betageometric lifetime process</td>
</tr>
<tr>
<td>( \Theta )</td>
<td>model-specific parameter vector</td>
</tr>
<tr>
<td>( \tau )</td>
<td>(unobserved) lifetime of a customer</td>
</tr>
<tr>
<td>( X(t) )</td>
<td>random variable for the number of purchases in the time period from ( T ) to ( T+t )</td>
</tr>
<tr>
<td>( Y(T,T+t) )</td>
<td>random variable for the number of purchases in an interval of length ( t )</td>
</tr>
<tr>
<td>( P_{\text{Pois}} )</td>
<td>Poisson probability defined as ( P_{\text{Pois}}(X(t) = x) = \left( \frac{\lambda t}{x!} \right)^x )</td>
</tr>
<tr>
<td>( P_{\text{NBD}} )</td>
<td>neg. binomial prob. defined as ( P_{\text{NBD}}(X(t) = x) = \left( r \cdot \alpha \cdot t \right)^x )</td>
</tr>
<tr>
<td>( {v} )</td>
<td>Pochhammer’s symbol</td>
</tr>
<tr>
<td>( B(a,b) )</td>
<td>beta function</td>
</tr>
<tr>
<td>( _2 F_1 )</td>
<td>Gaussian hypergeometric series</td>
</tr>
</tbody>
</table>

Note: All probabilities are implicitly conditional on their respective parameters, i.e., \( \lambda \) and \( p \) on the individual level and \( r, \alpha, a, b \) on the aggregate level.
Table 2: Key expressions of integrated purchase-incidence/lifetime-duration models

<table>
<thead>
<tr>
<th>Managerial Issue</th>
<th>Expression</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Describe purchase behavior</td>
<td>$L(\Theta</td>
<td>x, t, T)$</td>
</tr>
<tr>
<td>Measure heterogeneity</td>
<td>$P(X(t) = x)$</td>
<td>Probability distribution of purchase frequencies (Section 2.3.2)</td>
</tr>
<tr>
<td>Measure purchase rates</td>
<td>$P(\tau &gt; T</td>
<td>x, t)$</td>
</tr>
<tr>
<td>Assess model adequacy</td>
<td>$E(\binom{X(t)}{x})$</td>
<td>Expected number of transactions in a given time period $t$ (Section 2.3.4)</td>
</tr>
<tr>
<td>Identify individuals who most likely represent active or inactive customers - churn prediction</td>
<td>$E(\binom{X(T,T+t)}{x,t,T})$</td>
<td>Expected number of transactions of a customer in the time period $(T, T+t]$ conditional on his/her purchase history (Section 2.3.5)</td>
</tr>
</tbody>
</table>

Note: The expressions are conditional on an appropriate parameter vector $\Theta$.

geometrically distributed dropout process (A3), the individual level likelihood function is therefore given by

$$L(\lambda)p(x, t, T) = (1 - p)^x \cdot P_{Poiss}(X(t) = x) \cdot P(X(T - t) = 0).$$  

This result can be rewritten as

$$L(\lambda)p(x, t, T) = p \cdot (1 - p)^x \cdot P_{Poiss}(X(t) = x) + (1 - p)^{x+1}. $$

It is known that the mixture of Poisson purchases with gamma heterogeneity leads to a negative binomial distribution (cf. Ehrenberg 1959 for an early reference or Wagner/Taudes 1987 for a comprehensive overview). Therefore, if we account for heterogeneity in purchase rates across customers (A2), the likelihood takes the form

$$L(r, \alpha, \beta)p(x, t, T) = p \cdot (1 - p)^x \cdot P_{NBD}(X(t) = x) + (1 - p)^{x+1}. $$

Making use of (A5), the likelihood of the CBG/NBD results by accounting for heterogeneity (A4) in the dropout probability $p$ across customers:

$$L(r, \alpha, \beta)p \cdot P_{NBD}(X(t) = x) + (1 - p)^{x+1}. $$

After simplifying, we arrive at

$$L(r, \alpha, \beta)p(X(t) = x) = \binom{b}{a} \cdot \binom{a + 1, b + 1}{a + b} \cdot P_{NBD}(X(t) = x) \cdot \binom{a + b}{a}. $$

This result is noteworthy for two reasons: 1) The close relationship to the NBD model is directly obvious. The likelihood for an individual customer with purchase history $(x, t, T)$ is the weighted sum of two negative binomial probabilities, corrected by the betageometric deflection process. The first term represents a customer who made $x$ transactions but dropped out after time $t$, and the second term represents a customer who also made $x$ transactions but is still active at time $T$. 2) The negative binomial distribution is included by default in most spreadsheet and statistics software packages, making the estimation procedure based on (A5) a simple task.

The overall likelihood function value for a cohort of $n$ customers results by multiplying the individual likelihoods (5) of each customer. As repeated multiplication of small numbers leads to reduced numerical precision, the computationally more efficient log-likelihood function is the preferred way of estimating the model’s parameters. The overall log-likelihood (LL) of $n$ customers is obtained by summing the logarithms of the individual likelihoods and follows directly from this result.

2.3.2. Probability Distribution of Purchase Frequencies

As was pointed out, (2) is based on individual level data $(x, t, T)$ which varies across the cohort. We now concentrate on the probability distribution of purchase frequencies, $P(X(t) = x)$, thereby aggregating over different values for $t$, i.e., recency aspects do not matter here. This theoretical distribution will be compared to the empirical tabulation by computing various goodness-of-fit measures.

As above, an individual client who purchased $x$ times up to time $t_c$ can either drop out or may still be active at time $T$:

- Dropout after $t_c$: Since purchase incidence and lifetime duration are assumed to be independent pro-
cesses in (A5), such a customer would have made at least x purchases (because of defection he/she had no chance for further repurchasing) and considering (A1) and the geometric defection (A3) we arrive at
\[ p \cdot (1 - p)^x \cdot P_{\text{geo}}(X(t) \geq x). \] (6)
(6) also holds for x = 0 since, in this case, it reduces to p, i.e., the dropout probability.
• Still active at time T: The customer missed (x + 1) chances to quit and purchased x times, thus
\[ (1 - p)^{x+1} \cdot P_{\text{geo}}(X(t) = x). \] (7)
We combine (6) and (7) and proceed analogously to the previous subsection: accounting for purchase rate heterogeneity (A2) as well as for dropout heterogeneity (A4), the probability distribution is found to be
\[ P(X(t) = x) = \frac{(b_x)}{(a + b_x)^{x+1}} \cdot \left( \frac{a}{b_x + x} \cdot P_{\text{NBD}}(X(t) \geq x) + P_{\text{NBD}}(X(T) = x) \right). \] (8)
Again, this result allows for an intuitive interpretation and may be evaluated by means of standard software.

2.3.3. Probability of a Customer Still Being Active

If customers are not bound to a company by contract, their lifetimes τ are usually not observed. Managers interested in tracking the total number of active customers or trying to identify candidates for churn prevention may use the probability that a customer is still active as a guideline for further action. Due to (A3), a customer who bought at time t₀ is only allowed to drop out after t₀ and considering principles of conditional probabilities we find:
\[ P(\tau > T | X(t) = x) = \frac{p}{P(X(T - t_0) = 0)}. \] (9)
By reducing the fraction with \((1 - p)^x \cdot P_{\text{geo}}(X(t) = x)\) and making use of (1) and (2), (9) can be rewritten as
\[ P(\tau > T | X(t) = x) = \frac{(1 - p)^x \cdot P_{\text{geo}}(X(T) = x) \cdot (t/T)^x}{L(a|X(t) = x)}. \] (10)
Finally, we account for heterogeneity, (A2) and (A4), consider independence (A5) and find, in accordance with (4),
\[ B(a,b + x + 1) \cdot P_{\text{NBD}}(X(T) = x) \cdot (t/T)^x \]
\[ P(\tau > T | X(t) = x) = \frac{B(a,b)}{L(r,\alpha, b, x)} \cdot \left( \frac{a}{b + x} \cdot (\alpha + T) + \frac{\alpha}{\alpha + T} \right)^{-1}. \] (11)
This result points to a notable difference between models based on a betagometric lifetime process and the continuous Pareto lifetime: if the customer bought in \(t_0 = T\), the \(V_{\text{NBD}}\) will result in a probability of exactly 1 that the customer is active (see SMC, (11), (12), and (13)). In our model, defection occurs immediately after purchases and at time zero only. Therefore, the probability of being active is always strictly smaller than 1. The same holds for the FHL model if \(x > 0\).

2.3.4. Expected Number of Transactions

In Section 2.3.2 the probability distribution of purchase frequencies is discussed. It may be used as a means to evaluate in-sample goodness-of-fit for the whole time span, thereby aggregating over time, i.e., different values of \(t\). Alternatively, one might be interested in the evolution of average purchase frequencies. Since the models discussed here all implicitly assume stationarity, the expected number of transactions increases monotonically with an increasing length of the time span. Nevertheless, in the case of short-term marketing activities, which do not violate the stationarity assumptions, it may be interesting to check the time series of the theoretical number of transactions against the observed data for, e.g., diagnosing the impact of certain marketing campaigns at the aggregate level (this line of reasoning follows the recommendation of, e.g., Aaker 1970). The expected number of transactions in the interval from zero to \(t\) is given by
\[ E(X(t)) = \frac{b}{a - 1} \cdot G(r,b,a,\alpha|a,t) \] (12)
with
\[ G(v_1, v_2, v_3, v_4|a,t) := \left[ 1 - \left( \frac{v_1}{v_2 + v_3} \right)^v_4 \right] \cdot F(v_1, v_2 + 1; v_3 + a, \frac{t}{v_2 + v_3 + 1}). \] (13)
We note that (12) is only defined for values of \(a \neq 1\) and we discuss this issue in detail in the Appendix.

2.3.5. Conditional Expected Number of Future Transactions

Stochastic models postulate behavioral assumptions at the individual level and are designed to describe individual consumer behavior. Accordingly, individual-level data is employed for calibration purposes. In addition, it is possible to calculate individual level measures, thereby making use of the well-established regression to the mean phenomenon, i.e., that “[…] we can better predict what the person will do next if we know not only what that person did before, but what other people did” (Greene 1982, p. 130). An example of such a measure is discussed in Section 2.3.3: the probability of a customer being active. Now we concentrate on predicting a client’s future purchases, \(E(Y(T,T + t)|x,t_0,T)\). The notation \(Y(T,T + t)\) is used to emphasize that we are dealing with ex-ante forecasts, starting at the end of the observed time span \(T\), \(t\) periods ahead. A analysis might be carried out at the individual level, but customers might also be grouped to form segments according to, e.g., frequency \((x)\) and recency \((T - t)\). Information of this kind is intended to assist marketing management decision making. In the case of the cbs/nbd, this measure is found to be
\[ E(Y(T,T + t)|x,t_0,T) = \frac{a + b + x}{a - 1} \cdot G(r,b,a,\alpha|a,t) \cdot P(\tau > T | x,t_0). \] (14)
Amongst others, the expected number of future transactions contains an expression which linearly depends on frequency and on the probability of this customer still being active. This, in turn, depends on recency. We note again that (14) is only defined for values of \(a \neq 1\) and discuss this issue in the Appendix.
3. Empirical Results

3.1. Replication Study

To provide empirical evidence for the performance of the CBG/NBD, we fit the model to a publicly available dataset. We employ the customer base used by FHL, replicate their results for the P/NBD and the BG/NBD, and add our model variant, the CBG/NBD. To make the results of estimations comparable between models, the likelihood formulation is based on the timing process for all three models. Because of this, similarities and differences can be pointed out clearly.

The parameters are estimated by the maximum likelihood method to the same fit period as FHL. A framework on top of the statistics environment R (R Development Core Team 2006) computes likelihood functions and further expressions for all three models [6]. The results of the calibration are displayed in Table 3. The standard errors (se) of the models’ parameters given are their asymptotic lower bounds according to the inverse of the information matrix (cf. Casella/Berger 1990, pp. 307 ff. and 325 ff.), the standard errors of \( r/\alpha \) and \( s/\beta \) and \( a/(a+b) \) are first-order Taylor approximations (cf. Casella/Berger 1990, pp. 328 ff.).

As this replication is not the focus of this article, we only briefly describe our findings. The maximum log-likelihood function values (LL) of the two BG/NBDs are almost identical (see Table 3) and larger than that of the P/NBD. All parameters are highly significant except for \( \beta \) of the P/NBD (however, larger standard errors for this scale parameter have been reported quite frequently in the literature, e.g. by Schmittlein/Peterson 1994 and Krafft 2002, pp. 106 ff.). Assumptions (A1) and (A2) are postulated for all three models, indicating that active customers will purchase \( r/\alpha \) times per period on average. The estimates in Table 3 suggest that, under the CBG/NBD, customers will purchase more often. (A3) and (A4) imply that after a purchase, customers will defect with a mean probability of \( a/(a+b) \). Following Table 3, dropout after purchase is less likely under the BG/NBD than under its variant. Taken together, the higher mean purchase rate, but also the higher mean drop-out probability, compensate each other.

When comparing the theoretical (cf. Section 2.3.2) and the observed tabulation, it is found that the BG/NBD matches the non-buyers class best; however, for the one-time buyers and the two-time buyers, the central version of the model comes closest (because of space limitations detailed results are not shown here). A formal \( \chi^2 \) goodness-of-fit test (see Table 3) indicates, that the CBG/NBD fits this dataset well; its value of the \( \chi^2 \)-statistic is the smallest. In the holdout period, the expected number of transactions, conditional on the number of transactions in the fit period (cf. Section 2.3.5), result in similar forecasts for both BG/NBD (no formal comparison was carried out since observed values are not available). Except for the one-time buyers, the CBG/NBD predicts slightly lower numbers of transactions.

3.2. An Application in Direct Marketing

3.2.1. Customer Base

Data Description

For our empirical study, we use a customer base provided by a medium-sized, German catalog retailer. The
company specializes in office supplies, ranging from low-priced copying paper and printer cartridges to premium computer and network components, and sells exclusively to business customers, who either buy for their own needs or resell. A excerpt from the commercial register is required before a new customer is served for the first time. Part of the customer base was made available for analysis consisting of those 1 875 customers that made their first purchase between March 4, 2001 and covering the time span up to March 3, 2003.

These data were used to create two non-overlapping sets: a dataset exclusively containing the 1 304 commercial end users (“Dataset 1”) and a dataset containing the 571 resellers (“Dataset 2”). In this study, we do not differentiate between the different product categories offered in the company’s catalog but rather account for all purchases. However, the distinction between the two segments of customers is considered important. According to managers familiar with this market, resellers are more loyal than end consumers. The product offering of the resellers remains relatively stable over a longer period. They made initial investments in their own marketing activities such as preparing printed advertising material and compiling price lists. Once acquired, this makes staying with their supplier (i.e., our retailer) for a longer duration in their own interest. In contrast, end consumers with few or no investments into the buyer-supplier-relationship may be easily tempted to switch suppliers responding to marketing activities of competitors. If the models considered here capture the customer behavior in these datasets correctly, these behavioral differences should emerge when comparing the various measures provided.

Summary Statistics

Both datasets are split into two periods: a fitting period and a holdout period. The fitting period ends on March 4, 2002; and the holdout period comprises the subsequent 52 weeks. As outlined in Section 2.1, we account for individual timing as we pool customers from different weeks. At the end of the fit period, the times of reference T of the customers range from 52 weeks for those who made their initial purchase on March 5, 2001 to as few as 43 weeks for those who purchased on May 6, 2001 for the first time.

Table 4 displays properties of both datasets. In both periods, the resellers have a higher mean number of purchases than the commercial end users. This originates from both the higher mean number of purchases per buyer and the higher proportion of buyers in that dataset.

3.2.2. Postulated Assumptions

We verify whether there is a sufficient degree of concordance between the behavioral assumptions of the models and the observed data [8]. The comparison in Table 4 of the descriptive statistics for two subsequent periods of time gives some indications on the conjecture of customers’ limited lifetimes and the implicit supposition of stationarity.

- A look at the proportion of buyers shows a decrease from the first period to the second. A formal test indicates that these changes are significant. This decrease makes it reasonable to assume that most of these customers have indeed defected. Another observation that can be made with regard to the proportion of buyers is that the relative decrease is stronger for Dataset 1. These differences are in line with the managerial assessment and should later be reflected in the results of the model calibration. We expect the models to indicate that the end consumers in Dataset 1 have lower purchase rates and shorter lifetimes than the resellers in Dataset 2.

- For both datasets, the decrease in the mean number of purchases of customers (i.e., all individuals in the cohort) is significant. This is in accordance with our expectations because customers are defecting; they do not purchase anymore but remain in the cohort. On the
Table 5: Parameter estimates and goodness-of-fit tests over a fit period of 52 weeks

<table>
<thead>
<tr>
<th></th>
<th>P/NBD</th>
<th>BG/NBD</th>
<th>CBG/NBD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dataset 1 (end users)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>LL</strong></td>
<td>-6001.94</td>
<td>-6011.65</td>
<td>-6006.91</td>
</tr>
<tr>
<td><strong>r (se)</strong></td>
<td>0.92 (0.13)</td>
<td>0.18 (0.01)</td>
<td>0.65 (0.07)</td>
</tr>
<tr>
<td><strong>a (se)</strong></td>
<td>14.96 (1.94)</td>
<td>2.56 (0.33)</td>
<td>5.57 (0.21)</td>
</tr>
<tr>
<td><strong>s (se)</strong></td>
<td>0.40 (0.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>β (se)</strong></td>
<td>2.00 (0.64)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>a (se)</strong></td>
<td></td>
<td>0.84 (0.18)</td>
<td>1.00 (0.14)</td>
</tr>
<tr>
<td><strong>b (se)</strong></td>
<td></td>
<td>1.81 (0.49)</td>
<td>1.10 (0.21)</td>
</tr>
<tr>
<td><strong>r / a (se)</strong></td>
<td>0.061 (0.004)</td>
<td>0.069 (0.006)</td>
<td>0.117 (0.010)</td>
</tr>
<tr>
<td><strong>s / β (se)</strong></td>
<td>0.204 (0.047)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>a / (a+b) (se)</strong></td>
<td>0.32 (0.023)</td>
<td>0.47 (0.039)</td>
<td></td>
</tr>
<tr>
<td>**P(τ &gt; T</td>
<td>x, t)**</td>
<td>0.27</td>
<td>0.76</td>
</tr>
<tr>
<td><strong>χ² (df)</strong></td>
<td>19.25 (6)</td>
<td>18.08 (6)</td>
<td>8.75 (6)</td>
</tr>
<tr>
<td><strong>p-value</strong></td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
<td>0.19</td>
</tr>
</tbody>
</table>

| **Dataset 2 (resellers)** | | | |
| **LL**  | -6088.78 | -6094.35 | -6089.04 |
| **r (se)**  | 0.72 (0.10) | 0.42 (0.03) | 0.86 (0.12) |
| **a (se)**  | 6.96 (0.86) | 4.14 (0.47) | 6.42 (0.69) |
| **s (se)**  | 0.18 (0.06) |  |  |
| **β (se)**  | 3.06 (2.84) |  |  |
| **a (se)**  |  | 0.23 (0.08) | 0.24 (0.04) |
| **b (se)**  |  | 1.96 (0.96) | 0.74 (0.21) |
| **r / a (se)**  | 0.104 (0.008) | 0.100 (0.008) | 0.134 (0.011) |
| **s / β (se)**  | 0.058 (0.037) |  |  |
| **a / (a+b) (se)**  | 0.10 (0.023) | 0.25 (0.042) |  |
| **P(τ > T | x, t)**  | 0.61 | 0.83 | 0.56 |
| **χ² (df)**  | 2.87 (6) | 7.71 (6) | 2.50 (6) |
| **p-value**  | 0.82 | 0.31 | 0.87 |

Note: LL stands for the value of the logarithm of the overall likelihood function, se for standard error, and df for degrees of freedom. The probability of not having defected is averaged across customers, yielding $P(\tau > T | x, t)$. 

other hand, the changes in the mean number of purchases of buyers (i.e., those who actually bought) are not significantly different between periods for both datasets. While this result is not sufficient to conclude that purchase rates of individual buyers are stationary, it gives a first indication that this might in fact be the case.

Next, we briefly discuss the appropriateness of the Poisson assumption and the corresponding supposition of exponentially distributed interpurchase times. Following the argumentation of Dunn/Reader/Wrigley (1983) and Wagner/Taudes (1986), a gamma distribution is fitted to the interpurchase times of each customer who bought at least five times. If the interpurchase times are distributed exponentially, the shape parameters $r$ of these gamma distributions should not be significantly larger than 1.

In order to be able to include as many customers as possible, we do not split the periods in this analysis, but rather include all observed purchases. By proceeding in this way, 158 customers remain in Dataset 1 and 208 in Dataset 2. For 5.1% and 11.5% of them, respectively, the hypothesis of exponentially distributed interpurchase times has to be rejected (see Tab. 4). Since these proportions are regarded as small, we conclude that Poisson type models are a reasonable choice for these data.
As lifetime-duration is not observed here (but must rather be inferred), the appropriateness of (A3), (A4), and (A5) cannot be tested beforehand but must be critically questioned when evaluating the results provided by the models. The latter also applies for (A2), the suitability of the gamma mixing distribution.

3.2.3. Model Calibration

The same software as before is used to calibrate the three models to the fit period of up to 52 weeks for both datasets. The \( \text{P/NBD} \) serves as our reference model, because it is the only model with multiple empirical applications in the literature (FH; Schmittlein/Peterson 1994; Kumar/Reinartz 2000; Wu/Chen 2000; Krafft 2002, pp. 91 ff.). Both \( \text{BCG/NBD} \)s have the role of the “challengers” (the plain \( \text{NBD} \) is calibrated as well for later use in the assessment of forecasting accuracy).

Parameter estimation is conducted in all cases by maximizing the overall log-likelihood function \( (LL) \) with a numerical search routine. The results are displayed in Table 5 and are discussed subsequently.

- **Logarithm of the overall likelihood function:**
  - All \( LL \) values are of similar magnitude. For both datasets, the \( LL \) of the \( \text{P/NBD} \) are the highest and those of the \( \text{BG/NBD} \) are the lowest; the \( \text{CBG/NBD} \) lies in between.

- **Parameter estimates:**
  - All parameters are of reasonable sizes and their standard errors are small (thus the estimates are highly significant) in all but one case: again, the scale parameter \( \beta \) (\( \text{P/NBD}, \) Dataset 2) is insignificant, and the shape parameter \( b \) (\( \text{BG/NBD}, \) Dataset 2) is weakly significant only. The gamma distributions describing purchase rate heterogeneity are monotonically decreasing \( r < 1 \), a pattern usually observed for such kinds of data. This also applies for heterogeneity of defection rates \( \text{P/NBD}, \) \( S < 1 \). The beta distributions describing heterogeneity of dropout probabilities are inverse-J-shaped in three cases and U-shaped for the \( \text{CBG/NBD} \) in Dataset 2. According to the polarization index, \( (a + b) \) is slightly more dispersed for the \( \text{CBG/NBD} \). The confidence intervals for shape parameter \( a \) for both \( \text{BG/NBD} \)s in Dataset 1 includes the “critical” value of 1 that no longer causes concern (cf. Appendix) – after rounding to three significant digits, a even equals 1.00 (\( \text{CBG/NBD}, \) Dataset 1).

- **Mean purchase rates \( r/x \):**
  - The fraction \( r/x \) indicates that all models attribute a higher expected number of purchases per period per active customer to the resellers than to the end consumers. A cross models, the \( \text{CBG/NBD} \) has the highest rate. This is a consequence of the models’ designs: since the \( \text{BG/NBD} \) does not allow non-buyers to defect, they have an attenuating effect on the average purchase rate. In contrast, the \( \text{CBG/NBD} \) allows these persons to become inactive immediately, leading to a higher average purchase rate of active customers.

These mean purchase rates might also be compared to mean(\( x \)) of customers/buyers as reported in Table 4. Thereby, one has to consider that results of Table 5 refer to a weekly analysis period while those of Table 4 refer to a yearly analysis period. Keeping the cohort pooling effects in mind as well as the drop-out process as postulated by the models, it becomes obvious that these results are not directly comparable. Nevertheless, as a rough yardstick, we divide mean(\( x \)) of buyers by 48 weeks, ending up with 0.058, 0.116 respectively. We concede that all models seem to overestimate the mean purchase rate for Dataset 1, the \( \text{CBG/NBD} \) markedly so. For Dataset 2, we observe some degree of underestimation \( \text{P/NBD}, \) \( \text{BG/NBD} \) and overestimation \( \text{CBG/NBD} \).

- **Customer defection:**
  - Conceptually, the most important difference between the models considered here is their assumption on the lifetime duration process; therefore the parameters \( s/\beta \) (mean defection rate) and \( a/(a + b) \) (mean dropout probability) are not directly comparable. Nevertheless, a similar decrease in magnitude for these figures as well as an increase of \( r/a \) is observed for all models when examined across datasets. This is in accordance with prior expectation because of the different behavioral patterns for end users and resellers.

The average probability that customers will not have defected by the end of the fit period, conditional on their purchase history, offers a means to compare retention across models (cf. Section 2.3.3) [9]. The difference in \( P(\tau > T|x,t) \) clearly mirrors the additional dropout possibility (with probability \( a/(a + b) \)) at time zero for the \( \text{CBG/NBD} \). We note that the results of \( \text{CBG/NBD} \) and \( \text{P/NBD} \) are very similar and in accordance with their approximations by the proportion of buyers as given in Table 4.

The \( \text{BG/NBD} \) seems to overestimate \( P(\tau > T|x,t) \) quite a bit. A discrepancy between \( \text{P/NBD} \) and \( \text{BG/NBD} \) with respect to this measure is also observed for \( \text{FH} \)’s data (cf. Table 3). The evidence is limited at this point but it might be interesting to analyze whether this finding extends to other datasets as well.

In summary, the parameter estimates are face valid and the models capture the observed real-world phenomena correctly. The measures arising from model calibration comply with the managerial assessment and the summary statistics discussed earlier (cf. Section 3.2.1).

3.2.4. Model Fit

The probability distribution of purchase frequencies (cf. Section 2.3.2) can be used to assess the fit of the models at the aggregate level. Figures 2a and 2b plot observed tabulations against the theoretical distributions. As can be seen from these graphs, no model reproduces the observed frequency distribution perfectly. Nevertheless, for both datasets, the \( \text{CBG/NBD} \) matches the observed fre-
quences of the two largest classes (non- and one-time-buyers) better than the other models. In contrast, the BG/NBD overstates the frequency of the one-time buyers in both cases, exhibiting spikes for this class. This visual inspection gives us some confidence that the CBG/NBD is more appropriate for these datasets than the original model.

This preliminary conjecture is supported by a formal \( \chi^2 \) goodness-of-fit test. We use classes \( x = 0...9 \) and a remainder class for \( x \geq 10 \). The key finding of this formal test is that the CBG/NBD dominates both the P/NBD as well as the BG/NBD for these data. The results are summarized in the last two rows for each dataset of Table 5.

### 3.2.5. Predictive Quality for Out-Of-Sample Forecasts

**Forecasting Accuracy Based on Relative Absolute Errors**

The expected number of future transactions is a central outcome of the models under discussion, forming an important building-block for the planning of future marketing management activities. In this section, the quality of predictions is assessed and a formal error measure is used, which is appropriate when there is interest in the accuracy of individual-level forecasts.

Evaluating forecast precision at the individual level is carried out by means of the relative absolute error (RAE). For the present case, this measure is defined as

\[
\text{RAE}_{T,T} = \frac{\mathbb{E}(Y(T,T + t)|x,T) - (x_{T+1} - x_T)}{\mathbb{E}_{\text{nbd}}(Y(T,T + t)|x,T) - (x_{T+1} - x_T)}.
\] (15)

In this equation:

\( \mathbb{E}(Y(T,T + t)|x,T) \) is the conditional expected number of future transactions as discussed in Section 2.3.5. We use the notation \( x_T \) in order to emphasize that this measure is based on the observations made in the fit period, i.e., purchases up to the time \( T \).

\( \mathbb{E}_{\text{nbd}}(Y(T,T + t)|x,T) \) is the conditional expected number of future transactions based on a plain NBD model. In this context, the NBD seems a natural choice of a baseline or naive model.

\( (x_{T+1} - x_T) \) is the actual number of purchases occurring only in the holdout period. Subtracting purchases observed by time \( T \) is necessary since \( \mathbb{E}(Y(T,T + t)|x,T) \) predicts purchases occurring after that time.

The relative absolute error can be interpreted as follows: a value of 0 implies a perfect forecast. Values between 0 and 1 indicate that predictions are an improvement over the naive forecast. Finally, values greater than 1 represent projections less accurate than the naive forecast. The NBD is applicable as a naive forecast in the sense that it is based on the assumption of an unlimited lifetime; \( \mathbb{E}_{\text{nbd}}(Y(T,T + t)|x,T) \) extrapolates the updated purchase rates of individuals linearly into the future, not accounting for defection.

Individual level quantities are averaged across customers by employing the geometric mean, thus resulting in Armstrong/Collopy’s (1992) geometric mean relative absolute error (GMRAE). The interpretation of this measure resembles that of the RAE. This measure is communicated easily, allows comparisons between different datasets, and was found to be reliable (ibid.).
Tashman (2000) recommends the use of different fit and forecasting periods to evaluate out-of-sample forecasting accuracy, so that conclusions will be less affected by particularities of the business environment in those intervals. In view of the available holdout sample of 52 weeks and a reasonable balance between the lengths of estimation and forecasting periods, we decided to use intervals containing 39 or 52 weeks. The previous sections reported results for an estimation period of 52 weeks only. When repeating the analysis for 39 weeks, there is no change to the basic findings described beforehand. Because of limitations in space and the focus of this paper being the comparison of BG/NBD models, Table 6 shows the GMRAEs for these two models only, for each combination of estimation and forecasting periods.

As expected, all values in Table 6 are much smaller than 1, implying improvements over the NBD forecasts. Thus, the lifetime assumptions in both models do indeed pay off. Table 6 reveals two rather consistent patterns that persist across the different lengths of fit and forecasting intervals. 1) The improvement of either model over the NBD is substantially smaller for Dataset 2 than it is for Dataset 1. This finding has an intuitive explanation. The information given by managers familiar with this market (cf. Section 3.2.1) and the estimation results from Section 3.2.3 point to lower defection rates of resellers. Since purchase process assumptions are the same (i.e., NBD), this higher stability in Dataset 2 is likely to result in behavioral patterns that are closer to the NBD baseline model. 2) While the BG/NBD forecasts show a considerable improvement over the NBD, the central version outperforms the BG/NBD for this customer base consistently with respect to forecasting accuracy. The figures provided in Table 6 also allow a direct comparison between the two BG/NBDs: since geometric means have been employed for averaging, one just has to divide two corresponding GMRAEs to arrive at a GMRAE with the denominator’s model as baseline, e.g., for Dataset 2, 52 weeks of estimation and forecasting period, 0.63/0.72=0.88 is the GMRAE of the CBG/NBD in comparison to the BG/NBD.

GMRAE is a relative measure of forecasting precision and it is shown that both BG/NBDs perform well in comparison to a plain NBD. In addition, they also perform well in terms of an absolute measure. We take the absolute errors (numerator of (15)), average over customers, and end up with MAD, the mean absolute deviation. For Dataset 1 we find MADS of 0.76, 0.75, and 0.73 for the P/NBD, the BG/NBD, and the CBG/NBD, i.e., the error encountered when forecasting the number of purchases an (average) individual customer will make within the next year is less than one purchase. For Dataset 2, these figures are 2.28, 2.28, and 2.24. To make these MADS comparable across datasets, we divide them by the observed numbers of purchases in the holdout period and again find a better prediction accuracy for Dataset 1.

Forecasting Accuracy Based on Comparison per Purchase Frequencies

In view of the motivation to relax the assumption of the FH model (cf. Section 2.2), an improvement over the BG/NBD is expected primarily for customers with low repeat after trial. Table 7 and Figures 3a and 3b analyze this conjecture more closely.
Based on RAES, we determined for each customer (in each combination of estimation and holdout periods) which of the two BG/NBDs forecasts more accurately. Table 7 is set up to pay further attention to whether this customer did not buy at all or just once (x ≤ 1), or did buy more than once (x > 1), in the forecasting period. Once more, the results are consistent across the combinations of different lengths of fit and prediction period: 1) The CBG/NBD forecasts dominate in the majority of the cases. 2) The advantage is especially profound for the no or low repeat-after-trial class (x ≤ 1). 3) This improvement is more pronounced for Dataset 1.

It was pointed out already that the conceptual cornerstones of our model are frequency and recency of past purchases. Whereas Table 7 evaluates the predictive accuracy by concentrating on the former aspect, we now focus on the latter [11]. More recent purchases increase clients’ probabilities of being active and, therefore, this information might also influence forecasting precision. Figures 3a and 3b are constructed based on the same information as before (i.e., which model yields the smaller RAЕ for each individual customer). The horizontal axes refer to the time of the last purchase, t, which occurred sometime in weeks 1 to 52. The vertical axes refer to the recency of this last purchase, i.e., the difference between T and t, (T - t) varies between 0 and 52 in the present context. For obvious reasons (T ≤ 52), not all combinations of (t, T - t) can occur, which results in the diagonal structures of Figures 3a and 3b. The width of these diagonal bands is directly related to the length of the interval used to build the customers’ cohorts (cf. Section 2.1). Forecasts for clients with similar (t, T - t) are grouped to form a common hexagon in Figures 3a and 3b. The pie charts within these hexagons indicate the percentage of times either model predicted better than the other. The area of the pie charts is proportional to the number of observations in the particular hexagon.

Both graphs support our conjecture: for customers who made their last repurchase early or do not repurchase at all (left upper parts), the CBG/NBD forecasts future behavior with a higher relative accuracy. On the other hand, there are cases where the BG/NBD provides better predictions, i.e., clients with very recent last purchases (right lower parts), especially for Dataset 2.

4. Conclusion

4.1. Discussion of the Results

This article considers a modification of a well-established model, FHl’s BG/NBD. This variant aims at increasing the model’s flexibility while simultaneously retaining its strengths. The motivation for doing so stems from frequently-observed patterns of consumer behavior, i.e., that many customers who try a certain product do not repurchase it. We show that by slightly adjusting the assumptions postulated by FHl, this trial/repeat phenomenon can be accounted for more closely.

The CBG/NBD, our alternative to FHl’s model, is applied to three different datasets. A brief replication study of the
data analyzed by FHL demonstrates that this variant results in reasonable parameter estimates and a slightly better fit. Its superiority with respect to accounting for trial/repeat (in particular for buyers with high recencies and low frequencies) is illustrated by investigating datasets on end-users and resellers, both being served by a German catalog retailer. This claim is verified by parameter estimates which are face valid and moreover statistically significant, by an appropriate in-sample fit and by improved individual ex-ante predictions (for different lengths of fit and holdout periods).

Buying patterns of end users seem to be even more suited for the model under consideration. Conceptually, the low involvement purchase decision processes of end users are very much in accordance with the general idea of stochastic models of consumer behavior and the postulated Poisson assumption appears to match very well. The more routinized purchase decision processes of retailers are also accurately described, though to a somewhat lesser extent. Nevertheless, the parameter estimates and the statistics provided reflect these differences in consumer behavior.

The CBG/NBD estimates higher purchase intensities as well as higher deflection probabilities when compared to the BG/NBD. These impacts compensate for each other on average. Comparing both models with the P/NBD, the BG/NBD is closer with respect to the purchase process, and the CBG/NBD with respect to the dropout process. However, all models exhibit the same tendencies.

In addition to these insights with respect to marketing issues, we hope this paper makes a contribution in model-building. The alternative approach taken to develop the model via counting processes results in relatively slim mathematical derivations (in comparison with the papers by SMC and FHL), analytical expressions which assist intuitive interpretation (especially within an NBD framework), and ease of software implementation (e.g., in Microsoft Excel or R). Moreover, we provide justification for an unrestricted domain for the coefficients of both BCg/NBD models, making it feasible to provide confidence intervals ubiquitously.

4.2. Limitations

From the marketing side, our conclusions lack support from a broader variety of different empirical datasets since only three have been analyzed so far. This implies that some of the conclusions provided above are tentative in character.

From the modelling side, four issues seem to be of particular concern to us: 1) Ignoring quantity selection: Since the random variable analyzed is the buying occasion, purchase quantity is ignored. This might be especially disturbing in markets where purchases of more than one product at a time occur frequently. However, e.g., Fader/Hardie/Lee (2005b), demonstrate the substantial increase necessary in model sophistication when trying to account for this aspect. 2) Assuming independence between purchase incidence and lifetime duration: SMC discuss this assumption extensively and find arguments both for and against it. Even when replacing this assumption, the type of dependence is by no means obvious and thus open to debate anyway. 3) Postulating a very simple timing process: The information employed per individual is reduced to \((x_{t_i},T)\), thus ignoring \((t_1,t_2,\ldots,t_{n_i-1})\), which is usually available to the model builder. These purchase incidents might include information on dynamic constituents of customer behavior (e.g., seasonality or learning). On the other hand, it is obvious that considering \((x_{t_i},T)\) only reduces the models' complexity substantially. 4) Omitting managerial controls and neglecting competition: Reducing the scope of analysis to a very limited perspective represents a characteristic feature of stochastic models of consumer behavior in general, and NBD-based models in particular. However, this reduces the variety of potential applications.

These limitations are moderated by the fact that the fundamental concept of FHL is practicability by elementariness. Overcoming them would require at least parts of this aspiring goal to be sacrificed.

4.3. Areas of Future Research

Of course, most of the issues raised above represent areas of future research. We add some other suggestions, which all aim at extending the analysis to markets with somewhat different types of purchase behavior in order to broaden/limit the applicability of the models discussed.

Section 2.2 provides examples of three stylized companies serving different types of customers. Company A, the Internet store, might be visited occasionally without customers placing an order. This kind of information could be used as an indication that such a patron is still active. Can these models process such an input? Company B, the classical retailer, mails catalogs to potential new clients. The date of this activity could be used as time zero for the BG/NBDS (instead of the date of the first purchase). What consequences result from doing so? Company C, the cellular phone service provider, has an almost contractual relationship with its customers. Is the Poisson assumption still a good choice for such a situation?

Future research might be performed along these lines in order to investigate whether/how these models can be extended to the different marketing environments outlined. Since we do not claim general superiority of either of these three models, such analysis should also generate recommendations as to which model is most appropriate under which market conditions.

Finally, a thorough simulation study might be performed analyzing the robustness of the models with respect to deviations from the basic assumptions. This could also result in a deeper understanding of the determinants that influence in-sample fit and forecasting accuracy. The
latter may be crucial for applying these models on a stan-
dardized basis in business.

Appendix – Removable Discontinuities

It was already mentioned that (12) and (14) are not defined for
\( a = 1 \) as the denominator becomes zero. A closer examination,
removes that this problem is inherent to both models since it
also occurs for \( \text{NBD} \)'s model (cf. \( \text{FHL} \), (9) and (20)). Investigating
the numerators of (12) and (14), we find that they also approach
zero, since \( f_1 \) reduces to an elementary function (Abramowitz/
Stegun 1974, §1.1.5):

\[
\lim_{n \to 1} G(v, y, z, v_1, v_2, t) = 1 - (v_1 v_2 t)^{-\gamma} (1 - \frac{1}{v_1})^{-v_1} = 0.
\]

As a whole, we are dealing with an indeterminate form.

Is this a relevant issue? In empirical applications, the chances of
considering a model with a value of exactly \( a = 1 \) are negligible.
However, this consideration is incomplete. Strictly speaking, the
estimate of a must be significantly different from 1, typically in
terms of a confidence interval (12). Otherwise, the model-builder
might run into problems of methodological soundness. The
parameter \( a \) affects the shape of the beta distribution describing
heterogeneity of dropout probabilities across customers. In
marketing applications, this distribution is frequently U- or inverse-J-
shaped, since there are usually two segments of either high core
loyals (p near zero in this case) or customers with a very high
dropout propensity (p near 1), implying \( a < 1 \). Thus, in practice
one might indeed find confidence intervals including the critical
value of 1. This also applies for the work of \( \text{FHL} \), who do not report
standard errors in their paper. Our replication study on their data-
set (see Section 3.1 and Table 3), however, indicates that both
their original model and our central version result in estimates of \( a \)
that are not significantly different from 1.

Therefore, we analyzed the behavior of (12) and (14) when
approaches 1 more closely. We did not experience difficulties in
numerical terms, since the variations in \( E(X(t)) \) and \( E(Y(T, T +
t)|x, T) \) are small and change smoothly. It is shown now in three
steps that \( \lim_{n \to 1} E(X(t)) \) and \( \lim_{n \to 1} E(Y(T, T + t)|x, T) \) exist and are finite.

Step 1: Review of standard results
- The derivative of the Pochhammer symbol is defined as (Abramow-
itz/Stegun 1974, §6.3.5)

\[
d_\alpha \Gamma(x) = (x - \alpha) \Gamma(x - \alpha)
\]

with \( \Delta_\alpha \Psi(x) := \Psi(x + \alpha) - \Psi(x) - \Psi(\alpha) \) the digamma func-
tion.
- The derivative of \( 1/(x)_\alpha \) follows to be

\[
d_\alpha \Gamma(1/x) = - \Delta_\alpha \Psi(x)
\]

The Gaussian hypergeometric series is defined as

\[
_2F_1(a, b; c; z) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n n!} \frac{z^n}{n!}
\]

with \( \partial_{a} _2F_1(a, b; c; z) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n n!} \frac{\Delta_n \Psi(c)}{\Gamma(c)} \frac{z^n}{n!} \)

Step 2: Application of l'Hôpital's rule
With \( \alpha \), numerator and denominator of (12) approach zero in
the limit. Hence, l'Hôpital's rule (Abramowitz/Stegun 1974,
§3.4.1) gives us

\[
\lim_{n \to 1} E(X(t)) = \lim_{n \to 1} \frac{\partial}{\partial a} G(v, y, z, v_1, v_2, t)
\]

\[
= \frac{b}{1 - a} \text{const.} \cdot \lim_{n \to 1} \frac{\partial}{\partial a} _2F_1(v_1, v_2 + 1, v_2 + 1, a; z)
\]

for \( v_1 = r, v_2 = b, v_2 = a \) and \( z = t(v_1 + t) \). We use (20):

\[
\lim_{n \to 1} \frac{\partial}{\partial \alpha} _2F_1(v_1, v_2 + 1, v_2 + 1, a; z) = - \lim_{n \to 1} \frac{\partial}{\partial \alpha} \sum_{n=0}^{\infty} \frac{(v_1)_n (v_2 + 1)_n}{(v_2 + 1)_n n!} z^n
\]

with

\[
\zeta_a = \frac{(v_1)_n (v_2 + 1)_n}{(v_2 + 1)_n n!} \Delta_\alpha \Psi(v_2 + 1, a).
\]

Step 3: Convergence of the power series (22)
This power series is convergent for \( z \) within its circle of conver-
gence \( \rho \) (Heuser 2006, §63.1). Since \( z \) is strictly less than 1, the
finite limit of (12) exists for \( \rho = 1 \). If \( \xi < 0 \) for (almost) all terms
of the series, the circle of convergence \( \rho \) can be determined by:

\[
\rho = \lim_{n \to \infty} \frac{\xi (n + 1)}{\xi (n + 1)} = \lim_{n \to \infty} \frac{\Delta_\alpha \Psi(v_2 + 1, a)}{\Delta_\alpha \Psi(v_2 + 1, a)}
\]

(23)

The sandwich theorem for sequences (Heuser 2006, §22.2) is used
for calculating \( \rho \). In order to apply this theorem, we construct two
sequences \( \eta^a \) and \( \eta^b \), which minorise and majorise \( \xi \), respecti-
ely, i.e., \( \eta^a < \xi < \eta^b \) for almost all \( n \). If we find \( \eta^a \) and \( \eta^b \) which
match this condition, (b) have the same limit and (c) the limit is 1,
then \( \rho = 1 \).

- Sequence \( \eta^a \):

\[
\Delta_\alpha \Psi(v_2 + 1, a) < \xi (n + 1) \Rightarrow \Delta_\alpha \Psi(v_2 + 1, a) < \xi (n + 1)
\]

Thus,

\[
\lim_{n \to \infty} \eta^a = \lim_{n \to \infty} \frac{(n + 1)(n + v_2 + a)}{\xi (n + 1)} = \frac{n}{\xi (n + 1)}
\]

(24)

- Sequence \( \eta^b \):

\[
\Delta_\alpha \Psi(v_2 + 1, a) < \xi (n + 1) \Rightarrow \Delta_\alpha \Psi(v_2 + 1, a) < \xi (n + 1)
\]

Thus,

\[
\lim_{n \to \infty} \eta^b = \lim_{n \to \infty} \frac{(n + 1)(n + v_2 + a)}{\xi (n + 1)} = \frac{n}{\xi (n + 1)}
\]

(25)

It follows that \( \rho = 1, 21 \) exists and is finite and therefore \( E(X(t)) \)
(12) may be extended continuously for \( a = 1 \): the discontinuity is
removable.

The existence of \( \lim_{n \to 1} E(Y(t, T + t)|x, T) \) follows accordingly: with
(16), the numerator and denominator of (14) also approach zero.
Therefore, we again apply l'Hôpital's rule:

\[
\lim_{n \to 1} E(Y(T, T + t)|x, T) = \lim_{n \to 1} \frac{\partial}{\partial a} G(v_1, v_2, v_2 + 1, v_2, t)\]

(25)

The denominator of (25) is defined and continuous in a neighbor-
hood of \( a = 1 \) and the limit of \( \frac{\partial}{\partial a} G(v_1, v_2, v_2 + 1, v_2, t) \) was shown to
exist previously (cf. Step 3).

This proof can be extended to the \( \text{FHL} \) model as well.

Notes
[1] The abbreviation \( \text{SMC} \) will be used for Schmittlein/Morrison/Colombo (1987) from here on.
[2] The abbreviation \( \text{FHL} \) will be used for Fader/Hardie/Lee (2005a) from here on.
In a strict sense, the term "purchase occasion" would be more appropriate since the aspect of quantity selection is ignored.

In order to avoid excessive notational burden, we avoid customer-specific indices throughout the paper, although \( (x_i, T) \) are all measured at the individual level.

Thus, the importance of precisely accounting for pooling increases with the length of the pooled period and should always be considered carefully.

The software package provides routines for the NBD, \( \text{pNBD} \), \( \text{BG/NBD} \) and the \( \text{CBG/NBD} \) model. Maximum likelihood calibration employs the R stats4 library. The model expressions and the computation of the hypergeometric series use the \( \text{RMP} \) multi-precision library to warrant accurate numerical results.

The corresponding expression for the \( \text{pNBD} \) model is given by \( \text{SNC, (11)}-(13) \), and a result for the \( \text{BG/NBD} \) results from \( \text{FHL} \), (12).

Except for the class \( x = 9 \) in Dataset 2, where the observed number of nine-time-buyers equals 4, all observed and predicted frequencies are greater than five, thereby meeting the requirements for a \( \chi^2 \) test (Cochran 1954).

Since different lengths of estimation and prediction period seem to have no substantial influence on forecasting quality, and because of space limitations, we will restrict the analysis henceforth to the 52/52 scenario.

We thank Albert Bemmaor for pointing this out.

References


