Small Sample Properties of the Pareto/Negative Binomial Distribution Model

by Daniel Hoppe and Udo Wagner

A gap remains between the development of marketing models by academics and their application by practitioners. Scholars identified thorough validation and broad understanding of a model's potential and limitations as central requirements for its widespread use. This article therefore provides a specific analysis of the Pareto/Negative binomial distribution model, the standard approach to describing and predicting consumers' long-term behavior in noncontractual settings. A literature review indicates that the empirical applications of this model remain scarce; therefore, this study offers results pertaining to its sampling properties from a Monte Carlo simulation study with an emphasis on small sample sizes. In so doing, the authors revise existing application recommendations by considering the characteristics of the particular customer data under analysis, including releasing some of the restrictions on sample size requirements and the number of observation periods.

Keywords
Customer lifetime value, purchase behavior, simulation, stochastic models of consumer behavior

1. Introduction

The availability of marketing data from corporate purchase records, panels, and surveys, and the increasing demands of managers willing these data to support their decision making represent key drivers of the maturation of marketing models (Leefflang/Wittink 2000). Opportunities for standardized applications and automated decision making arise from models with wide ranges of possible uses and minimal conceptual flexibility to prevent discouraging practitioners. For example, the Pareto/Negative binomial distribution model (from hereon Pareto/NBD model) (Schmittlein/Morrison/Colombo 1987) combines a broad range of possible uses with widespread research support; using simple assumptions and parsimonious data requirements, the model can answer a range of managerially relevant questions. In particular, it is well-suited for tracking customer equity, making automated mailing decisions, predicting churn, operationalizing customer lifetimes, and forecasting customer lifetime values in businesses that lack contractual ties. With this broad spectrum of applications, the Pareto/NBD model represents an excellent candidate for a standard model in customer-base analysis.

However, to exploit a model, practitioners demand guidelines for its application in their specific marketing context. For standard methodologies such as multivariate regression analysis or principal component analysis, vast literature gives precise recommendations about the best practices and requirements (e.g., Fox 1997). The basic NBD model (without explicit consideration of consumer lifetimes) similarly has been applied to various empirical settings and analyzed extensively (e.g., Ehrenberg 1988; Chatfield/Goodhardt 1973; Frisbie 1980; Schmittlein/Bemmaor/Morrison 1985). The Pareto/NBD extension can address customer defection. Despite the importance of such a model, with its integrated lifetime process, no guidelines exist for its application in different marketing environments. To resolve this problem, this study offers an extensive Monte Carlo simulation that analyzes the model's properties, especially among small samples. As this model is applied to company-internal cohorts of new customers as opposed to consumer panels, the findings therefore are highly relevant for small and medium corporations with low customer acquisition rates.

The remainder of this study proceeds as follows: Section 2 briefly reviews the theoretical properties of the Pareto/NBD model and provides an overview of its prior empirical applications. The gaps in extant knowledge
about this model provoke the research questions in Section 3. These research questions then guide the design of a Monte Carlo simulation, as introduced in Section 4. Section 5 contains the results of the simulation, and Section 6 concludes with a discussion of the findings.

2. Pareto/NBD Model

2.1. Postulated Behavioral Assumptions and Data Requirements

The Pareto/NBD model derives from five straightforward assumptions (see Schmittlein/Morrison/Colombo 1987):

A1 Individual purchase incidents by customers follow a Poisson process with rate $\lambda$.

A2 Each individual customer lifetime follows an exponential distribution with rate $\mu$. Customer-base analysis often features the concept of a defection rate ($DR$) within unit time. Exponentially distributed lifetimes correspond to a constant $DR$ equal to $[1 – \exp(-\mu)]^{-1}$.

A3 The individual-level purchase rates $\lambda$ are gamma distributed across customers with the shape parameter $r(> 0)$ and the scale parameter $\alpha(> 0)$, such that $E(\lambda) = r, \alpha$, and the coefficient of variation $CV_\lambda = 1/\sqrt{r}$.

A4 The individual-level lifetime rates $\mu$ are gamma distributed across customers with the shape parameter $s(> 0)$ and the scale parameter $\beta(> 0)$, such that $E(\mu) = s, \beta$, and $CV_\mu = 1/\sqrt{s}$.

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

The assumption of constant individual-level purchase rates $\lambda$ and lifetime rates $\mu$ implies stationary conditions. The mixture of Poisson-distributed purchases with gamma heterogeneity yields the well-known NBD model. Compounding an exponential lifetime distribution with gamma heterogeneity results in a Pareto distribution of the second kind (Johnson/Kotz/Balakrishnan 1994), and the combination of both processes constitutes the Pareto/NBD model. This model is parsimonious in terms of the number of parameters (i.e., four) but demanding with regard to its mathematical and computational complexity [2].

To apply the resulting model, managers must turn to cohort analysis which groups together customers who experience a common event at the same time. In this context, the common event is the beginning of the customers’ lifetime with the company, usually indicated by a patron’s first purchase from a certain supplier. After defining suitable cohorts, in order to estimate the parameters ($r, s, \alpha, \beta$) of the Pareto/NBD model, minimal customer-specific information is required [3]:

$x$ is the number of purchases a customer makes in a given time span (excluding a possible purchase at time zero).

$t_f$ is the time of the customer’s last purchase, that is, the time of purchase incident $x$. This value must be strictly positive, though for notational convenience $t_f = 0$ if no purchases are observed for this customer.

$T$ is the time of reference.

2.2. Empirical Applications

The Pareto/NBD model is the starting point for a class of stochastic models exhibiting attractive properties for the continuous management of a customer base (e.g., Hoppe/Wagner 2007 provide an overview on typical applications). However, the initial research effort pertaining to the Pareto/NBD model (Schmittlein/Morrison/Colombo 1987) included only illustrative examples, it was not applied empirically for several years. In 1987, the computational demands for model calibration and the evaluation of measures and statistics implied by the model placed a high burden on researchers and practitioners interested in its applications. The need to supply precise information about customer acquisition and purchase incidents from previous years rendered an application impractical in the era of floppy disks and 16-bit processors. Several years later though, Schmittlein/Peterson (1994) applied the model through a two-step, method-of-moments procedure, which they used to calibrate the model among catalog retailer customers and predict the individual-level reorder amounts. Kumar/Reinartz (2000) and Krafft (2002) also have used the model to assess customer lifetime duration in the catalog retailing industry. Reinartz/Kumar (2003) extend this idea to their concept of profititable lifetime duration, and Wu/Chen (2000), Fader/Hardie/Lee (2005a), Batislam/Denizel/Filiztekin (2007), Hoppe/Wagner (2007), Jérab/Fader/Hardie (2009) use the Pareto/NBD model as a benchmark for their newly developed models. Fader/Hardie/Lee (2005b) also have derived an expression for expected discounted residual transactions, thus forging a link to customer lifetime value.


Yet in contrast with the widespread impact of this model on research, the number of empirical applications remains quite limited. Moreover, of the publications that apply the model, several rely on the same data set. Publications in leading marketing journals that use empirical
parameters of the lifetime process differences across studies. Average lifetime rate, provide rough yardsticks for com-

bining substantial heterogeneity. The granularity spans 12 to 78 units (months or weeks). This granularity from 73 to as many as 4,576 customers, and the time spans 12 to 78 units (months or weeks). This granularity difference makes parameter estimates incomparable, though \( \frac{DR}{yr} \) year, which is the average number of purchases per year by an active customer, and \( \frac{DR}{month} \), or the monthly defection rate of an active customer with an average lifetime rate, provide rough yardsticks for comparisons across studies.

The estimates for the NBD part of the model, \( r \) and \( \alpha \), indicate mixing distributions with inner modes (\( r > 1 \)) and therefore moderate heterogeneity, \( CV_r < 1 \), as well as monotonously decreasing shapes (\( r \leq 1 \)), which indicate substantial heterogeneity. Tab. 1 exhibits a smaller range of average yearly purchases by an active customer (\( \frac{PR}{year} \)) compared with the range of average monthly defection rates (\( \frac{DR}{month} \)).

### 3. Research Questions

Winer (2000) argues that the penetration of marketing models into business practice has been generally overstated. Faced with a decision about whether their sparse resources should be devoted to understanding and implementing a marketing model, practitioners must consider several questions: “Is this model right for our company? Will this model capture the behavior of our customers sufficiently well?” Therefore, researchers must pose the research question: “Are there properties of customer bases that indicate a priori if the Pareto/NBD model can and should be applied?” Although the authors of the studies in Tab. 1 claim that the Pareto/NBD model possesses a high degree of face validity, they do not address its applicability in practice in sufficient detail. Moreover, the Pareto/NBD model has been applied to a rather limited number of data sets, product categories, and industries thus far, which prevents a formal meta-analysis (Hanssens et al. 2009; Leeflang et al. 2000). The only available guidelines with regard to the adequacy of the Pareto/NBD model come from Schmittlein/Peterson (1994). For one cohort of their customer base, they analyze stability (i.e., widths of calibrated confidence intervals) in the expected number of transactions by an active customer who exhibits an average purchase rate against two characteristics: the cohort size \( (N) \) and the fit period length \( (T) \). On the basis of the bootstrap confidence intervals, they establish lower bounds for \( N \) and \( T \) of 1,600 customers and 12 months, respectively [4]. To extend this finding, this study broadens the scope of the analysis substantially.

A central concept for evaluating model performance involves the quality of all the model parameter estimates [5] (and their related statistics \( E(\lambda) = r/\alpha \) and \( E(\mu) = s/\beta \)). Casella/Berger (1990) recommend using the mean squared error (MSE) as a performance measure of parameter estimators (\( \hat{\theta} \)), because it can be decomposed into bias and variance components, such that \( MSE(\hat{\theta}) = [bias(\hat{\theta})]^2 + [sd(\hat{\theta})]^2 \). Therefore, the quality of the parameter estimates depends on the MSE and its two components. Regarding the potential characteristics of a customer base, this study considers two different influ-

### Table 1: Empirical Applications of the Pareto/NBD Model

<table>
<thead>
<tr>
<th>Study</th>
<th>Analyzed market</th>
<th>( N )</th>
<th>( T )</th>
<th>Scale</th>
<th>Estimation</th>
<th>( r )</th>
<th>( \alpha )</th>
<th>( \frac{PR}{year} )</th>
<th>( CV_r )</th>
<th>( s )</th>
<th>( \beta )</th>
<th>( \frac{DR}{month} )</th>
<th>( CV \lambda )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoppe/Wagner, Small Sample Properties of the Pareto/Negative Binomial Distribution Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schmittlein/Peterson (1994)</td>
<td>Office supply, B2B, USA</td>
<td>4,050 30 m MM</td>
<td>3.70</td>
<td>2.05</td>
<td>MM</td>
<td>4.05</td>
<td>12 m MM</td>
<td>3.51</td>
<td>4.72</td>
<td>.85</td>
<td>21</td>
<td>.03</td>
<td>1.00</td>
</tr>
<tr>
<td>Leeflang et al. (1994)</td>
<td>Catalog retailer, B2C, Germany</td>
<td>9,167 18 m MM</td>
<td>4.00</td>
<td>1.53</td>
<td>MM</td>
<td>9,167</td>
<td>12 m MM</td>
<td>4.00</td>
<td>1.53</td>
<td>.85</td>
<td>21</td>
<td>.03</td>
<td>1.00</td>
</tr>
<tr>
<td>Winer (2000)</td>
<td>Catalog retailer, B2C, USA</td>
<td>9,467 24 m MM</td>
<td>14.59</td>
<td>2.78</td>
<td>MM</td>
<td>9,467</td>
<td>52 m MM</td>
<td>14.59</td>
<td>2.78</td>
<td>.47</td>
<td>30</td>
<td>.96</td>
<td>1.83</td>
</tr>
<tr>
<td>Krafft (2002)</td>
<td>Catalog retailer, B2C, Germany</td>
<td>9,467 24 m MM</td>
<td>14.59</td>
<td>2.78</td>
<td>MM</td>
<td>9,467</td>
<td>52 m MM</td>
<td>14.59</td>
<td>2.78</td>
<td>.47</td>
<td>30</td>
<td>.96</td>
<td>1.83</td>
</tr>
<tr>
<td>Fader et al. (2005a)</td>
<td>Compact disc retailer, Internet</td>
<td>2,557 39 w MM</td>
<td>10.58</td>
<td>1.35</td>
<td>MM</td>
<td>2,557</td>
<td>78 w MM</td>
<td>10.58</td>
<td>1.35</td>
<td>.61</td>
<td>11.67</td>
<td>12.28</td>
<td></td>
</tr>
<tr>
<td>Battelmuss et al. (2007)</td>
<td>Grocery retailer, B2C, Turkey</td>
<td>5,479 52 w MM</td>
<td>4.04</td>
<td>1.33</td>
<td>MM</td>
<td>5,479</td>
<td>52 w MM</td>
<td>4.04</td>
<td>1.33</td>
<td>.47</td>
<td>4.06</td>
<td>.71</td>
<td></td>
</tr>
<tr>
<td>Jerath et al. (2009)</td>
<td>Grocery retailer, B2C, Turkey</td>
<td>5,479 52 w MM</td>
<td>4.04</td>
<td>1.33</td>
<td>MM</td>
<td>5,479</td>
<td>52 w MM</td>
<td>4.04</td>
<td>1.33</td>
<td>.47</td>
<td>4.06</td>
<td>.71</td>
<td></td>
</tr>
<tr>
<td>Hoppe/Wagner (2007)</td>
<td>Office supply, B2B, Germany</td>
<td>1,304 52 w MM</td>
<td>14.96</td>
<td>2.95</td>
<td>MM</td>
<td>1,304</td>
<td>52 w MM</td>
<td>14.96</td>
<td>2.95</td>
<td>1.04</td>
<td>.40</td>
<td>2.00</td>
<td>.72</td>
</tr>
<tr>
<td>Wu/Chen (2000)</td>
<td>Tea specialty store</td>
<td>1,366 48 w MLE</td>
<td>15.14</td>
<td>2.57</td>
<td>MM</td>
<td>1,366</td>
<td>78 w MLE</td>
<td>15.14</td>
<td>2.57</td>
<td>.56</td>
<td>4.33</td>
<td>.75</td>
<td></td>
</tr>
<tr>
<td>Fader et al. (2005a)</td>
<td>Compact disc retailer, Internet</td>
<td>2,557 39 w MLE</td>
<td>10.58</td>
<td>1.35</td>
<td>MM</td>
<td>2,557</td>
<td>78 w MLE</td>
<td>10.58</td>
<td>1.35</td>
<td>.61</td>
<td>11.67</td>
<td>12.28</td>
<td></td>
</tr>
</tbody>
</table>

Note: \( N \) cohort size, \( T \) fit period length, Scale time units are months (“m”) or weeks (“w”), Estimation method-of-moments (“MM”), and maximum likelihood (“MLE”).

Data sets, as displayed in Tab. 1, reveal some key consistent characteristics, such as cohort size \( N \) and the fit period length \( T \) to calibrate the model. Cohort sizes range from 73 to as many as 4,967 customers, and the time spans 12 to 78 units (months or weeks). This granularity difference makes parameter estimates incomparable, though \( \frac{PR}{year} \) year, which is the average number of purchases per year by an active customer, and \( \frac{DR}{month} \), or the monthly defection rate of an active customer with an average lifetime rate, provide rough yardsticks for comparisons across studies.

The estimates for the NBD part of the model, \( r \) and \( \alpha \), indicate mixing distributions with inner modes (\( r > 1 \)) and therefore moderate heterogeneity, \( CV_r < 1 \), as well as monotonously decreasing shapes (\( r \leq 1 \)), which indicate substantial heterogeneity. Tab. 1 exhibits a smaller range of average yearly purchases by an active customer (\( \frac{PR}{year} \)) compared with the range of average monthly defection rates (\( \frac{DR}{month} \)).
Operational and behavioral characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohort size</td>
</tr>
<tr>
<td>Fit period length</td>
</tr>
<tr>
<td>Purchase frequency</td>
</tr>
<tr>
<td>Lifetime duration</td>
</tr>
<tr>
<td>Purchase process heterogeneity</td>
</tr>
<tr>
<td>Lifetime process heterogeneity</td>
</tr>
</tbody>
</table>

Quality of all parameter estimates

MSE: mean squared error (bias² + variance)

Research questions

RQ1 Cohort size ↑ → MSE ↓
RQ2 Fit period length ↑ → MSE ↓
RQ3 Fit period length ↑ → MSE improvement more pronounced for lifetime than for purchase process
RQ4 Purchase frequency ↑ → MSE ↓
RQ5 Lifetime duration ↑ → MSE_{PR} ↓, MSE_{DR} ↑
RQ6 Tendency to underestimate amount of heterogeneity
RQ7 Effect sizes of characteristics
RQ8 Recommendations on cohort sizes and fit period lengths

Figure 1: Schematic Representation of Research Questions

ences: (1) two operational characteristics pertaining to the market under consideration, cohort size and fit period length (Schmittlein/Peterson 1994) and (2) four behavioral characteristics that follow directly from the model assumptions A1–A4, that is, how often customers purchase, how long they stay with the company, and how much they differ. The general design of the study, as displayed in Fig. 1, features the following research questions.

RQ 1 Does the quality of the parameter estimates of the Pareto/NBD model improve with moderately larger cohort sizes? According to statistical theory, maximum likelihood estimates are asymptotically unbiased. Therefore, bias and the MSEs for all four parameters should decrease with increasing cohort size, ceteris paribus. However, the quality resulting from moderate cohort sizes is not clear a priori.

RQ 2 Does the quality of the parameter estimates improve with a longer fit period? Observations over a longer time span provide more information about customer behavior. Although marketers generally want to implement a decision support tool as soon as possible, short fit periods may deteriorate the quality of the results (Fader/Hardie 2001b). Some evidence (Schmittlein/Peterson 1994) indicates that a short time span might result in overestimated customer defection rates, because the model might erroneously consider some active customers (especially those with low purchase rates who have not yet repurchased) already defected. Therefore, the MSEs of the parameter estimates should decrease with increasing fit period length, ceteris paribus.

RQ 3 Is the influence of fit period length on the quality of the estimates of the lifetime process parameters greater than on the quality of the purchase process parameters? General consensus indicates that the lifetime process entails less reliable estimates (e.g., Krafft 2002). Even short fit periods provide information about customers’ purchase processes. However, for customers with small purchase rates, short fit periods cannot capture the full purchase-repurchase cycle, so they may be wrongly designated inactive, which biases the parameter estimate of the lifetime process. Longer fit periods are more likely to capture long interpurchase times. Therefore, with a longer fit period, the reduction of MSEs should be more pronounced for parameter estimates of the lifetime process than for parameter estimates of the purchase process, ceteris paribus.

RQ 4 Does the quality of the parameter estimates improve for samples with higher purchase rates? The Pareto/NBD model inherently contains a trade-off between the probability of active customers who experience long censored spells and the probability of customer defection. Customers with high purchase rates indicate their active status more frequently and make this trade-off easier. Therefore, the MSEs for all four parameters should decrease with increasing purchase rates, ceteris paribus.

RQ 5 Does the quality of the parameter estimates depend on lifetime durations? Low defection rates provide more information about the purchase patterns of customers, whereas high defection rates for a given fit period indicate the lifetime distribution more clearly. The longer is the last censored spell, the less likely the customer is to be active for a given purchase rate. Therefore, MSEs for the two parameters that correspond to the purchase process should decrease with lower defection rates, whereas MSEs for the two parameters that correspond to the lifetime duration process should increase with lower defection rates, ceteris paribus.

RQ 6 Is the model capable of accounting sufficiently for heterogeneity in the purchase and the lifetime process? The degree of heterogeneity in a customer.
base is an important driver of segmentation, such as when marketers plan new product and service launches (Von Hippel 2005). In terms of the Pareto/NBD model, the shape parameters $r$ and $s$ describe heterogeneity and should reveal insights into the customer base. When they are limited in size, cohorts likely omit cases that stem from the extreme tails of the compounding distribution. Accordingly, variation in these cohorts might be lower than variation in the underlying population, and heterogeneity may be underestimated [6].

**RQ 7 Are the effects on the estimation quality of the characteristics of similar magnitudes?** Many characteristics of a customer base influence the model calibration, though the sizes of this influence likely vary. In contrast with Schmittlein/Peterson (1994), who concentrated on the effects of cohort size and fit period length for a given data set, the current study extends the scope of analysis to behavioral characteristics. However, there is no a priori indication about which influence should be more pronounced.

**RQ 8 Can sample size recommendations be reduced depending on the properties of the customer base?** Intuitive reasoning suggests a potential trade-off between the operational characteristics of the cohort size and fit period length against the behavioral characteristic purchase rates (e.g., the model might be more adequate for small samples if purchase rates are high).

Tackling these research questions requires knowledge of the true characteristics of the analyzed customer bases, which demands a Monte Carlo simulation study.

4. Experimental Setting

4.1. Experimental Scenarios

Scenarios suitable for a simulation study should reflect the most important characteristics of the corresponding real-world phenomena. As outlined in Section 3 and Fig. 1, the experimental scenarios are based on operational and behavioral characteristics. Because multiple other potential influences exist, relevant insights may stem from analyses of various misspecifications, including the correlation between individual-level traits, varying degrees of regularity in purchasing behavior, and non-stationarity. Yet in other simulation studies, researchers have argued that a comprehensive, systematic inclusion of all possible influences is hardly possible, because the number of factor combinations would grow at a combinatorial rate. Wagner/Reisinger/Gausteer (2001) conclude that a limitation to specific core aspects and the exclusion of misspecification can support a thorough examination of the capabilities of the model. Therefore, this study leaves the analysis of the effects of other potential influences for further research.

4.1.1. Operational Characteristics

Considering prior results and this study’s focus on the model’s applicability in small and medium corporations, the simulation uses cohort sizes $N$ of 250, 500, 750, 1,000, 1,250, and 1,500. For the fit periods, the literature review (cf. Tab. 1) indicates that estimating the defection process requires a minimal time span. To examine this effect systematically, this study includes fit periods lengths $T$ of 12, 18, 24, and 30 time units.

4.1.2. Behavioral Characteristics

The derivation of suitable ranges for the behavioral stimuli reflects the results of the literature review (cf. Tab. 1), which indicate substantial fluctuations in the mean purchase and mean defection rates, as well as wide variations across customers. The representative parameterization of behavioral characteristics begins by allocating these characteristics into ordinal categories. The verbal descriptions then can be transformed into measures, which are amenable to intuitive interpretations (e.g., by practitioners). Finally, the measures get converted into Pareto/NBD model parameter values for the data generation procedure.

1. First, for the ordinal categories, to avoid an excessive number of factor combinations, this study defines three categories for all behavioral characteristics: low, medium, and high.

2. Second, the transformation into business measures consists of four areas (see Tab. 2):

   - **Purchase frequency:** The $PF$/year column in Tab. 1 reveals values between two and eight. According to Chatfield/Goodhardt (1973), the range also may extend up to twelve purchases per year, equal to approximately one purchase per period (month), which corresponds with A1’s implication of interpurchase times with a mode at 0. The levels included in the simulation are $1/6$, $1/2$, and $1/1$ purchases per active customer per time unit.

   - **Defection rate:** The $DR$/month column in Tab. 1 yields average defection rates in the range .05 to 1.00. The extremely high defection rates observed by Schmittlein/Peterson (1994) result for short fit periods, which are unlikely to capture the purchase-repurchase cycles of customers with low

<table>
<thead>
<tr>
<th>Behavioral characteristic</th>
<th>Business measure</th>
<th>Ordinal level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>low</td>
</tr>
<tr>
<td>1. Purchase frequency</td>
<td>$\bar{v}(x)=$</td>
<td>$1/6$</td>
</tr>
<tr>
<td>2. Defection rate</td>
<td>$DR(\mu)=$</td>
<td>50/1000</td>
</tr>
<tr>
<td>3. Purchase process heterogeneity</td>
<td>$CV, X=$</td>
<td>$1/2$</td>
</tr>
<tr>
<td>4. Lifetime process heterogeneity</td>
<td>$CV, T=$</td>
<td>$3/4$</td>
</tr>
</tbody>
</table>

**Table 2: Parameterization of Behavioral Characteristics for the Experimental Scenarios**
purchase rates. For longer fit periods, these authors obtain a measure of slightly greater than .5. Studies that deal with similar issues but use different methodologies report attrition rates ranging from approximately .02 to .20 (e.g., Colgate/ Stewart/Kinsella 1996; East et al. 1998). Accordingly, the simulation herein uses values of .05, .275, and .5.

- **Purchase process heterogeneity:** The shape parameter of the gamma density relates directly to the coefficient of variation (CV) of this distribution (see A3 and A4) and therefore is well suited to characterize the heterogeneities inherent to purchase and lifetime processes. Schmittlein/Morrisson/Cooper (1993) consider shape parameters less than 1.0 very heterogeneous and those greater than 4.0 very homogeneous. Therefore, a CV of more than 1.0 describes a very heterogeneous population, whereas a CV smaller than .5 reflects a very homogeneous one. For this classification, the observed purchase processes in Tab. 1 (column CVp) cover the full range of populations; values of 1/2, 3/4, and 3/2 provide reasonable choices for the coefficient of variation of the purchase process.

- **Lifetime process heterogeneity:** Heterogeneity in lifetime rates is established similarly. Tab. 1 (column CVL) reveals only highly heterogeneous populations, though it seems reasonable to include more homogeneous cohorts in the simulation as well to account for the effects of other segmentation efforts. Therefore, this study uses values of 3/4, 1/1, and 2/1.

(3) Third, the conversion to Pareto/NBD model parameter values involves recalling the relationships between the parameters of the gamma distribution and the mean purchase rate, defection rate, and CV (i.e., A2–A4, respectively). The results indicate that \( r = CV_p^2, s = CV_p, \alpha = [E(\lambda) \cdot CV_p]^\gamma, \) and \( \beta = -[(n-1-DR(\mu)) \cdot CV_p]^\gamma. \) Behavioral characteristics, as defined in Tab. 2, thus get converted into Pareto/NBD model parameter values for the data generation for the experimental scenarios.

### 4.2. Experimental Procedure

With \( 3^4 = 81 \) combinations of behavioral characteristics, six cohort sizes, and four fit period lengths, the simulation involves 1,944 distinct characteristics. These scenarios encompass a wide range of characteristics of real-world customer bases. For each of the 1,944 scenarios, a random number and pseudo-cohort generation process created 100 data sets, as described by Hoppe (2007). The calibration of the Pareto/NBD model is rather tedious (Fader/Hardie 2005), because it requires repeated evaluations of Gaussian hypergeometric series. As preliminary studies indicated accuracy issues with this function in normal double precision arithmetics, all critical mathematical expressions were implemented using the high precision library MPFR (Fousse et al. 2007), which ensures stable evaluations of the likelihood function [7]. These algorithms can be embedded in a framework in the statistics environment R (R Development Core Team 2006). The software package offers convenient computation of the likelihood function and further measures of the Pareto/NBD model (see Hoppe 2007 for details). In order to ensure that the optimization routine would identify the global maximum of the likelihood function in most cases, the optimization was started from different values and the optimizer was restarted at the best parameters found so far (Press et al. 1992).

### 5. Analysis of Simulation Results

#### 5.1. Generation of Summary Statistics

The generated data can be structured in a matrix with 1,944 rows (Monte Carlo scenarios) and 6 columns (i.e., true \( \theta = (r,s,\alpha,\beta; r,l,\alpha,\beta) \)) [8]. For each scenario and each of the 100 replications, the generated individual purchase records, \((x,t,T)_n, n = 1,...,N\) provide the data sets to estimate the parameters \( \hat{\theta}_n, k = 1,...,100 \) by means of the ML-principle. In accordance with the performance measure \( \text{MSE}(\hat{\theta}) \) [9], as well as its decomposition into bias and variance components, another 18 columns join the data matrix to reflect the following three performance measures for each of the six parameters \((\hat{\theta}_n, l = 1,...,6)\):

1. Percentage bias \( \text{pbias}(\hat{\theta}) = \frac{\hat{\theta} - \theta}{\theta} \cdot 100 \),
2. Absolute percentage bias \( \text{abias}(\hat{\theta}) = \frac{|\hat{\theta} - \theta|}{\theta} \cdot 100 \),
3. Percentage standard deviation \( \text{psd}(\hat{\theta}) = \frac{s_d(\hat{\theta})}{\theta} \cdot 100 \),

where \( \hat{\theta} = \sum_{t=1}^{100} \hat{\theta}_n/100 \) and \( s_d(\hat{\theta}) = \sqrt{\sum_{t=1}^{100} (\hat{\theta}_n - \hat{\theta})^2}/99 \) are calculated over 100 Monte Carlo replications for all scenarios. As these three measures are percentages, the results are comparable across scenarios.

#### 5.2. Analysis of Descriptive Measures

Tab. 3 exhibits the summary statistics for \( \text{pbias}(\hat{\theta}), \text{abias}(\hat{\theta}), \) and \( \text{psd}(\hat{\theta}) \), indicating upward biased estimates on average, though the bias is quite small for the purchase process. Because the biases are measured as percentages, they appear almost negligible (e.g., for \( r/l,\alpha \)).

The latter measure reveals that the variability of the estimates (across replications) is substantial, especially for the lifetime process. The medians are smaller than the mean, which indicates skewed distributions, as also emphasized by some high maximum values that appear...
Table 3: Simulation Summary Statistics for Percentage and Absolute Percentage Bias and Percentage Standard Deviation

<table>
<thead>
<tr>
<th>( \hat{\theta}_i )</th>
<th>( \hat{r} )</th>
<th>( \hat{a} )</th>
<th>( \hat{r}/\hat{a} )</th>
<th>( \hat{d} )</th>
<th>( \hat{\beta} )</th>
<th>( \hat{s}/\hat{\beta} )</th>
<th>( \hat{p} )</th>
<th>( \hat{a}/\hat{p} )</th>
<th>( \hat{r}/\hat{a} )</th>
<th>( \hat{d} )</th>
<th>( \hat{\beta} )</th>
<th>( \hat{s}/\hat{\beta} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>-2.8</td>
<td>-2.0</td>
<td>-2.4</td>
<td>-29.9</td>
<td>-34.2</td>
<td>-3.3</td>
<td>4.0</td>
<td>4.4</td>
<td>1.4</td>
<td>5.5</td>
<td>12.2</td>
<td>5.2</td>
</tr>
<tr>
<td>Median</td>
<td>2.3</td>
<td>2.2</td>
<td>1.1</td>
<td>3.1</td>
<td>7.3</td>
<td>2.6</td>
<td>14.0</td>
<td>13.7</td>
<td>5.1</td>
<td>22.5</td>
<td>44.5</td>
<td>18.5</td>
</tr>
<tr>
<td>Mean</td>
<td>5.9</td>
<td>5.5</td>
<td>3.0</td>
<td>11.9</td>
<td>23.5</td>
<td>20.0</td>
<td>25.7</td>
<td>24.1</td>
<td>6.1</td>
<td>86.5</td>
<td>144.6</td>
<td>118.0</td>
</tr>
<tr>
<td>Maximum</td>
<td>129</td>
<td>112</td>
<td>9.4</td>
<td>761</td>
<td>1382</td>
<td>4606</td>
<td>584</td>
<td>576</td>
<td>30.0</td>
<td>7723</td>
<td>13967</td>
<td>38297</td>
</tr>
</tbody>
</table>

Note: Minima for \( abias(\hat{\theta}_i) \) are 0 in all cases; maxima are equal to the values for \( pbias(\hat{\theta}_i) \).

Reading example 1: In the scenarios where \( T = 1 \), \( pbias(s/\hat{\beta}) \) is roughly 50%. For the same scenarios, \( abias(s/\hat{\beta}) \) is close to 0%.

Reading example 2: The comparison of bias for different \( DR \) reveals that the higher \( DR \), the lower is the bias in \( s \). For \( r \), the adverse effect is observed, and the values are generally lower.

Figure 2: Mean Marginal Percentage Bias per Characteristic Level

Note: -, o, and + correspond to low, medium, and high levels of behavioral characteristics, respectively (\( PR \) – purchase rate, \( DR \) – defection rate, \( CV_P \) – coefficient of variation purchase rate, \( CV_L \) – coefficient of variation lifetime rate).

to represent outliers that affect the lifetime process particularly strongly. A more detailed analysis indicates that such outliers typically arise when the ratio of the parameters \( \hat{s}/\hat{\beta} \) lies on a slightly increasing ridge of the likelihood function [10], such as when the mean lifetime rate can be identified, but the shape of the heterogeneity distribution cannot. (The following sections show that these outliers primarily affect the smallest cohort size included in the simulation.) Moreover, the results are consistent with Schmittlein/Peterson’s (1994) results (see their figure 2). For \( T > 12 \), they find that the width of the bootstrap confidence interval for \( \hat{r}/\hat{\alpha} \) is approximately .075.

Assuming they used the 95% confidence level, their percentage standard deviation should be about 7.5 (for their estimate of the mean \( \hat{r}/\hat{\alpha} = 0.25 \)). This value coincides with the magnitude found herein (i.e., 6.1).

The statistics in Tab. 3 reflect the entire sample and do not relate to the different simulated scenarios. To capture the effects of the experimental settings, this study averages \( pbias(\hat{\theta}_i) \) separately for the different levels of all characteristics (e.g., for a cohort size of \( N = 250 \), \( pbias(\hat{\theta}_i) \) is the mean over all 324 corresponding scenarios). Fig. 2 depicts the findings: For each Pareto/NBBD parameter and influence level, a shaded square offers a
simple indication of the corresponding mean marginal bias – marginal in the sense that the information about the other influences has been averaged. In this graph, a darker shading indicates greater biases, and the values range from 0% to 80%.

Fig. 2 also provides some tentative responses to the research questions. Perhaps most important, the bias varies across the different levels of the characteristics, indicating that the characteristics chosen in Section 4, both operational and behavioral, are relevant. The estimates of the purchasing process are affected less by bias than the estimates of the lifetime process. The mean of the purchasing process (\( \hat{\theta} / \bar{\theta} \)) can be captured precisely across all levels of the characteristics. In contrast, the mean of the lifetime process (\( \bar{\theta} / \hat{\theta} \)) is captured to a much smaller extent, and the bias decreases with sample size \( N \) for all Pareto/NBD parameters. This finding is in line with the conjecture formulated by RQ1. For this study, it is worth emphasizing that the reduction in the mean percentage bias is especially high for the step from \( N = 250 \) to \( N = 500 \) customers. In agreement with RQ2, increasing the numbers of fit periods reduces bias, and the change is more pronounced for the lifetime process estimates than for the purchase process estimates, which responds to RQ3. Furthermore, Fig. 2 supports the implication of RQ4, because bias decreases with increasing purchase rates (PR). The comparison of the changes for different defection rates (DR) also supports the effect conjectured in RQ5; high defection rates reduce bias in the lifetime process, but they increase bias in the purchase process (though in the step from medium to high defection rates, the bias in \( \hat{\theta} \) increases slightly but it remains lower than the value for low dropout). Consistent with Tab. 3, there is a positive bias on average, such that the parameter estimates are greater than the true values of the parameters. Because higher values of the shape parameters \( r \) and \( s \) indicate less heterogeneity, this observation is in line with RQ6.

5.3. Inferential Analysis of Descriptive Measures

The research questions from Section 3 also can be investigated through regression models, which assess the impact of the characteristics chosen in the simulation experiment according to two performance measures, absolute percentage bias \( abias(\hat{\theta}) \) and percentage standard deviation \( psd(\hat{\theta}) \), for each of the six parameters [11]. Because no specific hypothesis predicts the impact of these characteristics on the performance measures available, simple models are acceptable. First, these models note the main effects only, leaving the analysis of potential interaction effects for further research. Second, the impacts of characteristics are assumed to be essentially linear, except for cohort size. Third, scaling of impacts must be considered, such that the operational characteristics, which use a ratio scale, are directly included in the models, whereas including behavioral characteristics, which rely on an ordinal scale, requires the use of level-specific dummy variables (e.g., \( T_{\text{medium}} \) = 1 in scenarios for a medium purchase rate, \( T_{\text{medium}} = 0 \) otherwise). These assumptions result in the following model, which is linear in the response coefficients and can be estimated by ordinary least squares (OLS) [12]:

\[
abias(\hat{\theta}) = a_0 + a_1 \bar{N} + a_2 \bar{N}^2 + a_3 T + \sum_{(PR, DR, CI) \in \text{medium}} b_1 I,
\]

where:
- \( a_0 \) intercept term, baseline of comparison for behavioral indicators
- \( a_1 - a_3 \) response coefficients for operational variables
- \( b_1 \) response coefficients for behavioral variables
- \( \bar{N} = N/1000 \) rescaling of cohort sizes for ease of reporting

The analysis of the descriptive measures reveals that the data are affected by outliers, which results in positively skewed distributions for both \( abias(\hat{\theta}) \) and \( psd(\hat{\theta}) \). For such data, Box/Cox (1964) recommend a log-transformation of the dependent variable to satisfy the normality assumption of a linear model. Atkinson’s (1985) score test indicates the adequacy of this transformation for all 12 dependent variables; therefore, this transformation is executed. In addition, the analysis of the potential distorting effects of the outliers relies on robust regression, based on Huber’s (1981) M-estimator applied to the original variables. On the whole (with minor exceptions, as marked in Tab. 4), the results are similar in substance (Hoppe 2007). The evidence therefore indicates that the results in Tab. 4 are not distorted by outliers or single influential observations.

Tab. 4 summarizes the calibration results for 12 regression analyses. Only 4 of (144) regression coefficients are not statistically significant at the 5% level. Their signs indicate the direction of their influence on the dependent variable (e.g., \( a_1 < 0 \), such that biases of all Pareto/NBD parameters decrease with increasing fit period length). However, these coefficients do not reflect the magnitude of their influence. For this reason, Tab. 5 exhibits effect sizes \( \eta^2 \) of all characteristics on the two estimation quality measures (e.g., mean purchase rate \( PR \) explains 28% of the total variation in \( \ln[abias(\hat{\theta})] \) contributes 28 percentage points to the \( R^2 \) of the corresponding regression model; Olejnik/Algina 2003).

5.4. Findings

The results of Tab. 4 and 5 in turn can be interpreted according to the research questions from Section 3.

RQ 1 Does the quality of the parameter estimates of the Pareto/NBD model improve with moderately larger cohort sizes? Consistently, \( a_1 < 0 \) and \( a_2 > 0 \), which results in increasing quality measures with increasing cohort sizes \( N \) across the range
analyzed [13], though at a decreasing rate. An extrapolation of this effect beyond the range of \( N = 1,500 \) would be misleading.

**RQ 2** Does the quality of the parameter estimates improve with a longer fit period? The coefficients for the fit period length \((a_d)\) are all negative, which indicates increasing quality measures with increasing fit period length. According to the effect sizes, this improvement seems rather small.

**RQ 3** Is the influence of fit period length on the quality of the estimates of the lifetime process parameters greater than on the quality of the purchase process parameters? Tab. 5 documents that the effect sizes \( \eta^2 \) for the purchase process parameters are consistently smaller than those for the lifetime process.

**RQ 4** Does the quality of the parameter estimates improve for samples with higher purchase rates? The coefficients of the dummy variable regressors for the mean purchase rate are consistently negative, and their (absolute) magnitudes increase from the medium to the high scenario. Therefore, higher purchase rates lead to reduced bias, as well as lower variance in the parameter estimates.

**RQ 5** Does the quality of the parameter estimates depend on lifetime durations? The coefficients of the dummy variable regressors for the detection

---

### Table 4: Impact of Simulation Characteristics on Estimation Quality

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>( \ln \left[ \hat{\text{bias}} (\hat{\theta}) \right] )</th>
<th>( \ln \left[ \hat{\text{psd}} (\hat{\theta}) \right] )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient</td>
<td>( \hat{r} ) ( \hat{\alpha} ) ( r_f \alpha ) ( \hat{z} ) ( \hat{\beta} ) ( s_f \beta )</td>
<td>( \hat{r} ) ( \hat{\alpha} ) ( r_f \alpha ) ( \hat{z} ) ( \hat{\beta} ) ( s_f \beta )</td>
</tr>
<tr>
<td>( a_0 )</td>
<td>4.50 4.34 4.00 6.26 6.30 4.35</td>
<td>4.82 4.72 2.51 6.96 7.10 5.10</td>
</tr>
<tr>
<td>( a_1 )</td>
<td>-3.39 -3.51 -1.34 -3.32 -3.64 -3.12</td>
<td>-2.02 -2.01 -1.51 -2.28 -2.48 -2.38</td>
</tr>
<tr>
<td>( a_2 )</td>
<td>1.05 1.13 0.82 1.08 1.21 0.98</td>
<td>-0.66 -0.66 -0.47 -0.76 -0.84 -0.82</td>
</tr>
<tr>
<td>( a_3 )</td>
<td>-0.03 -0.02 -0.01 -0.05 -0.05 -0.05</td>
<td>-0.02 -0.01 -0.01 -0.04 -0.03 -0.03</td>
</tr>
<tr>
<td>( b_{\text{PR; medium}} )</td>
<td>-1.36 -1.25 -0.68 -0.97 -1.07 -1.20</td>
<td>-0.80 -0.68 -0.44 -0.66 -0.80 -0.82</td>
</tr>
<tr>
<td>( b_{\text{PR; high}} )</td>
<td>-1.88 -1.65 -0.95 -1.28 -1.47 -1.66</td>
<td>-1.11 -0.94 -0.59 -0.85 -1.04 -1.07</td>
</tr>
<tr>
<td>( b_{\text{CVL; medium}} )</td>
<td>0.58 0.53 0.17 -1.30 -0.77 -0.85</td>
<td>0.50 0.42 0.29 -1.20 -0.81 -0.19</td>
</tr>
<tr>
<td>( b_{\text{CVL; high}} )</td>
<td>1.02 0.92 0.36 -1.58 -0.82 -0.76</td>
<td>0.80 0.68 0.48 -1.38 -0.83 -0.06</td>
</tr>
<tr>
<td>( b_{\text{CVL; low}} )</td>
<td>-0.38 -0.40 0.28 0.05 0.09 0.28</td>
<td>-0.26 -0.27 0.26 0.00 0.05 0.16</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>-0.62 -0.59 0.72 0.40 0.55 1.11</td>
<td>-0.45 -0.43 0.74 0.21 0.37 0.67</td>
</tr>
<tr>
<td>( R_{\beta} )</td>
<td>-0.60 -0.60 0.15 -0.81 0.14 1.18</td>
<td>-0.35 -0.37 -0.17 -0.60 0.04 0.97</td>
</tr>
<tr>
<td>( R_{\beta} )</td>
<td>0.64 0.59 0.05 0.23 0.45 0.60</td>
<td>0.85 0.81 0.94 0.72 0.64 0.77</td>
</tr>
</tbody>
</table>

Note: The OLS estimates use \( n = 1944 \); negative coefficients are in grey-shaded cells.

1 All estimated F- and t-values are significant at the 5% level except those marked.

2 OLS coefficients and coefficients estimated by robust regression possess different signs.

### Table 5: Effect Sizes of Simulation Characteristics on Estimation Quality

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>( \ln \left[ \hat{\text{bias}} (\hat{\theta}) \right] )</th>
<th>( \ln \left[ \hat{\text{psd}} (\hat{\theta}) \right] )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient</td>
<td>( \hat{r} ) ( \hat{\alpha} ) ( r_f \alpha ) ( \hat{z} ) ( \hat{\beta} ) ( s_f \beta )</td>
<td>( \hat{r} ) ( \hat{\alpha} ) ( r_f \alpha ) ( \hat{z} ) ( \hat{\beta} ) ( s_f \beta )</td>
</tr>
<tr>
<td>( \eta^2 )</td>
<td>.21 .22 .07 .15 .20 .14</td>
<td>.23 .26 .29 .16 .21 .20</td>
</tr>
<tr>
<td>( N )</td>
<td>.02 .01 .00 .05 .04 .04</td>
<td>.02 .01 .01 .05 .04 .05</td>
</tr>
<tr>
<td>( T )</td>
<td>.28 .23 .09 .11 .17 .18</td>
<td>.34 .29 .20 .12 .20 .23</td>
</tr>
<tr>
<td>( PR )</td>
<td>.08 .07 .01 .18 .06 .05</td>
<td>.17 .14 .12 .33 .15 .01</td>
</tr>
<tr>
<td>( DR )</td>
<td>.03 .03 .05 .01 .03 .08</td>
<td>.05 .06 .30 .01 .03 .09</td>
</tr>
<tr>
<td>( CV_P )</td>
<td>.03 .03 .00 .04 .01 .10</td>
<td>.03 .04 .02 .05 .01 .20</td>
</tr>
<tr>
<td>( CV_L )</td>
<td>.03 .03 .00 .04 .01 .10</td>
<td>.03 .04 .02 .05 .01 .20</td>
</tr>
</tbody>
</table>

Note: Effect sizes of less than .05 appear in grey text.
rate are consistently positive for the purchase process parameters but negative for the lifetime process parameters. Effect sizes for the defection rate are smaller than those for the purchase rate, with the exception of the estimation quality of \( \hat{h} \). Overall, bias and variance in the purchase process parameters increase with increasing defection rates, whereas bias and variance of the lifetime process parameters decrease.

RQ 6 Is the model capable of accounting sufficiently for heterogeneity in the purchase and the lifetime process? The shape parameters \( \hat{r} \) and \( \hat{s} \) are inversely related to the coefficient of variation for the mixing gamma distribution that models heterogeneity. Tab. 3 and Fig. 2 indicate that \( \hat{r} \) and \( \hat{s} \) are overestimated in most cases; accordingly, heterogeneity is underestimated. As expected, this effect influences the estimation of the scale parameters \( \hat{\alpha} \) and \( \hat{\beta} \).

RQ 7 Are the effects on the estimation quality of the characteristics of similar magnitudes? Tab. 5 reveals major differences in the effect sizes of the different experimental stimuli. The effects of cohort size, purchase rate, and defection rate are substantial. The influence of fit period length is relatively small, probably a consequence of the chosen time spans. That is, in line with Schmittlein/Peterson (1994), the minimal fit period length is 12. The effect of heterogeneity is rather small in general but seems more pronounced for \( psd(\hat{r}/\hat{r}e) \) and \( psd(\hat{s}/\hat{\beta}) \).

RQ 8 Can sample size recommendations be reduced depending on the properties of the customer base? There is strong evidence that Schmittlein/Peterson’s (1994) recommendations can be refined. The purchase and defection rates should be considered when establishing minimum cohort size requirements, depending on the requested quality of the parameter estimates. Tab. 6 provides a tentative example, including confidence intervals for \( (\hat{r}/\hat{\alpha}) \) for all nine factor combinations of the two most important behavioral characteristics (purchase frequency and defection rate) and all levels of cohort size. The width of these confidence intervals decrease with cohort size, and the entries of Tab. 6 provide the value of \( N \) that results in a width of less than 10% of \( (\hat{r}/\hat{\alpha}) \) for a type-I error of 5%.

6. Conclusion
6.1. Discussion
The bias and variance of the maximum likelihood estimator for the Pareto/NBD model might fluctuate between single-digit percentage deviations from the true parameter values up to more than three-digit percentage point deviations in extreme cases. The magnitude of both depends heavily on the characteristics of the data set. Sample characteristics such as cohort size and fit period length drive estimator quality; however, recommendations based solely on these operational characteristics fall short of accounting for other important influences. The purchase and defection rates of a customer base also influence estimator quality and might compensate for small cohort sizes or short fit periods. Therefore, practitioners should keep the following guidelines in mind:

- Appreciate limitations. This simulation yields several data sets in which the lifetime process could not be identified because the heterogeneity distribution degenerated to a mass point (all members of the cohort had almost the same lifetime rate). With such data sets, researchers should abstain from makeshift solutions, such as restraining the parameter space a priori. Such identification problems indicate the model is not right for the job, whether due to a lack of correspondence between the model assumptions and actual behavior or because of insufficient information in the data set. By restraining the domain of the search space to some particular subset, researchers “enforce” certain results because they apply a model to a data set that does not fit [14]. In such cases, it may be worthwhile to use a more appropriate time scale or revert to alternative models, such as the simpler exponential/NBD model.

- Scaling matters. Even with large cohort sizes and long fit periods, a Pareto/NBD application might not be adequate if the purchase rate is too low to provide sufficient information for model calibration. However, the purchase rate also is a function of the time unit (e.g., a customer who buys six times a year versus every other month). It may be worthwhile to consider the design of a diagnostic measure that indicates the adequacy of the model according to the granularity of a given customer base. Such a measure should trade off between a smaller number of time periods that contain more rich information and a larger number of time periods with sparse information (e.g., if a time span of 12 months is not sufficient, will 48 weeks be?). This task is appropriate as a topic for further research.

<table>
<thead>
<tr>
<th>Defection rate DR</th>
<th>low</th>
<th>medium</th>
<th>high</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purchase frequency</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>low</td>
<td>1,000</td>
<td>&gt;1,500</td>
<td>&gt;1,500</td>
</tr>
<tr>
<td>medium</td>
<td>500</td>
<td>1,000</td>
<td>1,250</td>
</tr>
<tr>
<td>high</td>
<td>500</td>
<td>750</td>
<td>1,000</td>
</tr>
</tbody>
</table>

Table 6: Cohort Size Requirements (for estimating \( \hat{r}/\hat{\alpha} \) with a relative precision of 10% and a type-I error of 5%)
Small samples may be worth it. Assuming that the pre-
conditions from the previous steps have been met,
small samples of 250 to 500 customers may provide
reliable and sufficiently precise estimates. This simu-
lation contains scenarios with 250 customers for
which the bias in both $\hat{\eta}$ and $s/\hat{\beta}$ is less than 1%,
while the individual parameter bias results in values
between 1% and 17% with moderate standard errors.
Such results will be acceptable for many applications.

6.2. Limitations and Further Research

Although the results of this analysis deepen understand-
ing of the Pareto/NBD model, its limitations suggest
some paths for additional research. First, this study con-
centrate on the small sample properties of the Pareto/
NBD model, focusing on the applicability for small and
medium corporations with low customer acquisition
rates. An equally valid investigation might analyze larger
samples or even the asymptotic properties of the model.
An extension of this investigation with sufficiently large
cohorts could provide pertinent insights.

Second, Monte Carlo simulations often suffer from some
arbitrariness in the derivation of the experimental stim-
uli. Despite the great care taken to arrive at relevant,
applicable scenarios, in some circumstances, a stretch to
more extreme values, whether very low or very high,
might be helpful. Equally feasible is the demand for finer
granularity within the chosen domain. Pragmatic consid-
erations prevent an analysis of such values herein,
because the computational burden already was quite sig-
ificant. Although the intrapolation of the findings
across the range of scenarios covered in this study does
not impose conceptual difficulties, extrapolation beyond
that range may be misleading.

Third, robustness in the case of deviations from the
model assumptions may be relevant. This analysis takes
a bird’s-eye perspective and remains restricted to rather
simple regression models, neglecting non-linearities
(except for sample size) and interactions. Although
this approach matches the goal of deriving guidelines
that will encourage practitioners to apply this model, a
more detailed analysis that accounts for non-linearities
and interactions might reveal additional, valuable
insights.

Notes

[1] In mathematical terms, $DR$ is $P(t - 1 \leq \tau \leq t | \tau \geq t - 1)$, where $\tau$ denotes lifetime.

[2] For the resulting analytical expressions, the reader is referred to the references given in Tab. 1.

[3] To avoid an excessive notational burden, this investigation does not use customer-specific indices, though $(s, \tau, T)$ are all measured at the individual level.

[4] This recommendation might prevent small and medium-sized companies with fewer customers from employing the Pareto/NBD model. They might simply pool new customers over longer time intervals, but doing so can result in cohorts exposed to different marketing activities. If this cohort size

requirement can be relaxed, the increase in the number of
potential applications of the Pareto/NBD model could be
substantial.

As can be inferred from Tab. 1, methods-of-moments cali-
brations have been used more frequently probably because their
computational requirements are less demanding. However,
maximum likelihood estimates possess superior theoretical
properties and will, therefore, be used in the present study.

If the means of the compounding distributions are captured
correctly, it follows from A3 and A4 that the bias will be
reflected in the scale parameters $\alpha$ and $\beta$ as well.

These implementations, however, require very powerful
hardware to perform the extensive computations.

The ratios $r/\alpha$ and $s/\beta$ offer intuitive interpretations
of the mean purchase and mean lifetime rates; they also make the
results comparable with Schmittlein/Peterson’s (1994) analy-
sis. In order to emphasize that these ratios are estimated at
the scenario level the notation $\hat{\eta} \epsilon s/\hat{\beta}$ is used.

To avoid an excessive notational burden, no scenario-spe-
cific indices are reported, though all performance measures are
calculated for each of the 1,944 scenarios.

If $s$ and $\beta$ reach infinity (given $s/\beta = y$), the Pareto/NBD
model collapses into the nested exponential/NBD model.

Because of limitations in space and potential difficulties
when performing Box/Cox transformation the results for $phisti(B)$ are not reported.

To avoid excessive notational burden, the parameter- and
performance-specific indices are neglected.

The software developed for this research project avoids this
problem, because it yields highly precise results, even for far
off values of the model parameters. A numerical search pro-
cedure can maneuver the parameter space freely.

References

Atkinson, A.C. (1985): Plots, Transformations, and Regression:
An Introduction to Graphical Methods of Diagnostic Regres-
sion Analysis, Oxford.

tion and Comparison of Models for Customer Base Analysis,
in: International Journal of Research in Marketing, Vol. 24, No. 3,
p. 201–209.


Bolton, R.N./Bowman, D./Briggs, E./Kumar, V./Parasur-
Customer Assets. A Framework for Customer Asset Manage-

Box, G.E.P./Cox, D.R. (1964): An Analysis of Transformations, in:
211–252.

Casella, G./Berger, R.L. (1990): Statistical Inference, Belmont,


Colombo, R./Weina, J. (1999): A Stochastic RFM Model, in: Jour-

East, R./Harris, P./Lomax, W./Wilson, G./Hammond, K. (1998): Customer Defection From Supermarkets, in: Advances in Con-

Hoppe/Wagner, Small Sample Properties of the Pareto/Negative Binomial Distribution Model


